

Indiana Law Review



Volume 25 No. 1 1991

INDIANA UNIVERSITY

APR 1 5 1992

SCHOOL OF LAW-INDPLS.

LIBRARY

ARTICLES

- Transfers Prior to Marriage and the Uniform Probate Code's Redesigned Elective Share—Why the Partnership is Not Yet Complete**
Rena C. Seplowitz

- Police Power Absolutism and Nullifying the Free Exercise Clause: A Critique of *Oregon v. Smith***
John Delaney

- Winning Isn't Everything, It's the Only Thing.
Violence in Professional Sports: The Need for Federal Regulation and Criminal Sanctions**
Daniel R. Karon

NOTES

- CERCLA's Web of Liability Ensnares Secured Lenders:
The Scope and Application of CERCLA's Security Interest Exception**

- Statutory Control of DNA Fingerprinting in Indiana:
An Analysis of *Koske v. Townsend Engineering*: The Relationship Between the Open and Obvious Danger Rule and the Consumer Expectation Test**

- "Lawsuit For Sale": An Analysis of the Assignability of Legal Malpractice Claims**

BOOK REVIEW

- Narrative and the Legal Neighborhood**
Thomas Patrick Gannon
Jay A. Ziemer

Career Move.

No doubt you plan to start a successful legal career soon, so here's a tip:

Start reading *Law Week* now. While you're still in law school.

Why? Because top attorneys nationwide read *Law Week* every week, and as you know, it's never a bad idea to follow in the footsteps of the best. But more importantly—

Law Week gives you complete coverage of today's most important cases, with immediate notification of precedent-setting court decisions in every legal field, plus unparalleled Supreme Court reporting.

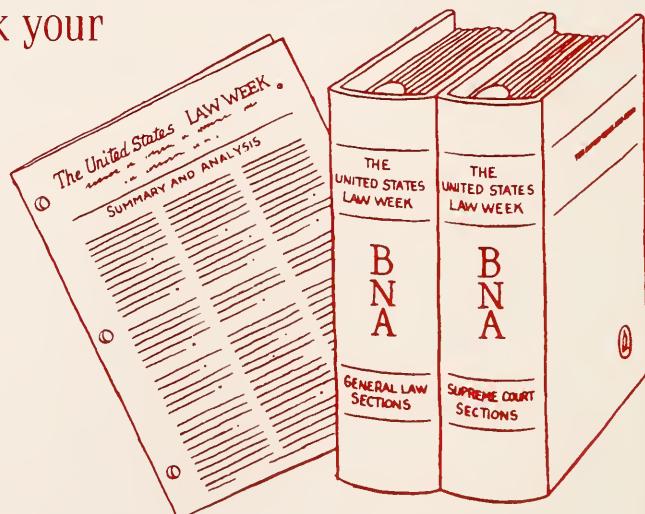
Indeed, *Law Week* puts you on the cutting edge of American law.

What better way to start a successful legal career.

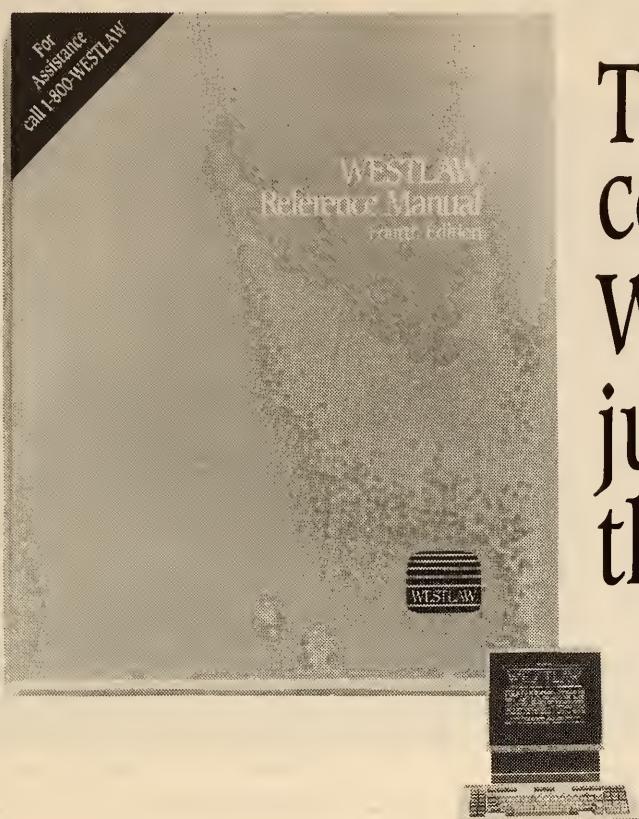
Make your move now. Ask your professor for details on *Law Week's* low student rates today.



The Bureau of National Affairs, Inc.
1231 25th Street, N.W.
Washington, D.C. 20037
202/452-4200



The United States
Law Week



To
command
Westlaw,
just crack
the book.

Commanding the power of WESTLAW® will benefit you now and throughout your legal career. And this new Reference Manual is the place to start.

THE ONE GUIDE AS NEW—AND NECESSARY—AS WESTLAW ITSELF.

The *WESTLAW Reference Manual, Fourth Edition* has been totally updated to help you take advantage of new WESTLAW enhancements and master your online research.

Over 500 pages, this inexpensive manual includes more than 230

quick-reference sheets describing individual databases and how to use them. PLUS... you'll discover the simplicity of EZ ACCESS, the timeliness of Shepard's PreView, the thoroughness of statutes research. Even how to investigate career opportunities on West's Legal Directory.

The reasons to use WESTLAW in law school and in practice are growing every day. The new *WESTLAW Reference Manual* helps you take advantage of them!

Available in your law school bookstore.



WEST PUBLISHING COMPANY

50 W. Kellogg Blvd. P.O. Box 64526 St. Paul, MN 55164-0526



**Please enter my subscription to the
INDIANA LAW REVIEW**

NAME _____

ADDRESS _____

Enclosed is \$_____ for _____ subscriptions.

Bill me for _____ subscriptions.

Mail to:

INDIANA LAW REVIEW
INDIANA UNIVERSITY
SCHOOL OF LAW—INDIANAPOLIS
735 West New York Street
Indianapolis, Indiana 46202

Subscription Rates (one year):

Regular, \$25; Foreign, \$28; Student, \$17 (4 issues)
Single Issue, \$8; Survey Issue \$17.00

Indiana Law Review

Volume 25

1991-92

Editor-in-Chief
Rolanda Moore Haycox

Executive Managing Editor
John A. Payton

Executive Editor

Articles
Jan S. Blaising

Executive Editor

Notes and Topics
Rebecca J. Seamands

Articles Editors

Scott R. Alexander
Frank C. Capozza
Linda J. Cooley
Mary T. Doherty
Joe Liebeschuetz
Jodie L. Miner
Courtney N. Stillman

Note Development Editors

Janet M. Charles
Rhonda L. Fuller
Katherine M. Gilson
James P. Leeds
Jerry L. Siefers, Jr.

Associate Editors

Joseph R. Hillebrand
Kathryn A. Hillebrands
Julia Bunton Jackson
Lawrence R. Kemm
Thomas E. Mixdorf
Laurie S. Morone
Rebekah N. Murphy

Heather K. Olinger
Stephaine L. Paine
Katherine Coble Pierce
Charles S. Roach
Vickie R. Shaw
Sally A. Voland
Barbara J. Weigel
Kathryn W. Williams

Editorial Assistant
Amy Morrison Grubbs

Faculty Advisor
Debra A. Falender

BGZ 0610

INDIANA LAW REVIEW

(ISSN 0090-4198)

Published four times a year by Indiana University. Editorial and Business Offices are located at:

Indiana Law Review
735 W. New York Street
Indianapolis, IN 46202
(317) 274-4440

Subscriptions. The current subscription rates are \$25.00 per four-issue volume (domestic mailing) and \$28.00 (foreign mailing). Unless the Business Office receives notice to the contrary, all subscriptions will be renewed automatically. Address changes must be received at least a month prior to publication to ensure prompt delivery and must include old and new addresses and the proper zip code.

Single and Back Issues. Current single regular issues may be obtained from the *Law Review* Business Office for \$8.00. Current Survey issues may be purchased for \$17.00. Please enclose payment with order. Complete sets and single issues of prior volumes are also available from the *Law Review* Business Office. Unless a claim is made for nonreceipt of issues within one year of the mailing date, we cannot guarantee supplying the issues without charge.

Manuscripts. The *Indiana Law Review* invites submission of unsolicited manuscripts. All citations should conform to *A Uniform System of Citation* (15th ed. 1991). Manuscripts should be triple-spaced with wide margins. The *Indiana Law Review*, as a matter of policy, promotes the use of gender-neutral language, and alternate use of the female or male pronoun is intended to refer to both female and male.

Copyright. All articles Copyright 1991 by The Trustees of Indiana University, except where otherwise indicated. All rights reserved. Opinions expressed are those of the contributor and are not presented as the views of the *Indiana Law Review*, its editors, or Indiana University School of Law-Indianapolis.

Indiana Law Review

Volume 25

1991

Number 1

Copyright © 1991 by the Trustees of Indiana University

TABLE OF CONTENTS

Articles

Transfers Prior to Marriage and the Uniform Probate Code's Redesigned Elective Share—Why the Partnership is Not Yet Complete	Rena C. Seplowitz	1
Police Power Absolutism and Nullifying the Free Exercise Clause: A Critique of <i>Oregon v. Smith</i>	John Delaney	71
Winning Isn't Everything, It's the Only Thing. Violence in Professional Sports: The Need for Federal Regulation and Criminal Sanctions	Daniel R. Karon	147

Notes

CERCLA's Web of Liability Ensnakes Secured Lenders: The Scope and Application of CERCLA's Security Interest Exception	Scott R. Alexander	165
Statutory Control of DNA Fingerprinting in Indiana	Joe Liebeschuetz	205
An Analysis of <i>Koske v. Townsend Engineering</i> : The Relationship Between the Open and Obvious Danger Rule and the Consumer Expectation Test	Jodie L. Miner	235
"Lawsuit For Sale": An Analysis of the Assignability of Legal Malpractice Claims.....	Rebecca J. Seamands	257

Book Review

Narrative and the Legal Neighborhood	Thomas Patrick Gannon	
	Jay A. Ziemer	289

Volume 25

1991

Number 1



The entire text of this Law Review is printed on recycled paper.

Indiana University School of Law—Indianapolis

1991-92 ADMINISTRATIVE OFFICERS AND FACULTY

Administrative Officers

THOMAS M. EHRLICH, *President of the University and Professor of Law.* LL.B., Harvard University.
GERALD L. BEPKO, *Vice-President, Indiana University; Chancellor, Indiana University-Purdue University at Indianapolis and Professor of Law.* J.D., IIT/Chicago-Kent College of Law; LL.M., Yale University.
NORMAN LEFSTEIN, *Dean and Professor of Law.* LL.B., University of Illinois; LL.M., Georgetown University.
THOMAS B. ALLINGTON, *Associate Dean for Academic Affairs and Professor of Law.* J.D., University of Nebraska; LL.M., New York University.
ARLENE G. ANDERSON, *Assistant Dean for Student Affairs.* J.D., Indiana University-Indianapolis.
JONNA M. KANE, *Assistant Dean for External Affairs.* J.D., Indiana University-Indianapolis.
JONI D. THOMPSON, *Assistant Dean for Development.* M.S., Ball State University.
ANGELA M. ESPADA, *Assistant Dean for Admissions.* J.D., Indiana University-Bloomington.

Faculty

THOMAS B. ALLINGTON, *Associate Dean and Professor of Law.* B.S., J.D., University of Nebraska; LL.M., New York University.
EDWARD P. ARCHER, *Professor of Law.* B.M.E., Rensselaer Polytechnic Institute; J.D., LL.M., Georgetown University.
JAMES F. BAILEY, III, *Professor of Law and Director of Law Library.* A.B., J.D., M.A.L.S., University of Michigan.
GERALD L. BEPKO, *Vice President, Chancellor and Professor of Law.* B.S., Northern Illinois University; J.D., IIT/Chicago-Kent College of Law; LL.M., Yale University.
ROBERT BROOKINS, *Associate Professor of Law.* B.S., University of Southern Florida; J.D., Ph.D., Cornell University.
DANIEL H. COLE, *Assistant Professor of Law.* A.B., Occidental College; A.M., University of Chicago; J.D., Lewis and Clark College; J.S.M., Stanford Law School.
PAUL N. COX, *Professor of Law.* B.S., Utah State University; J.D., University of Utah; LL.M., University of Virginia.
CLYDE HARRISON CROCKETT, *Professor of Law.* A.B., J.D., University of Texas; LL.M., University of London (The London School of Economics and Political Science).
DEBRA A. FALENDER, *Professor of Law.* A.B., Mount Holyoke College; J.D., Indiana University-Indianapolis.
DAVID A. FUNK, *Professor of Law.* A.B., College of Wooster; J.D., Case Western Reserve University; M.A., The Ohio State University; LL.M., Case Western Reserve University; LL.M., Columbia University.
PAUL J. GALANTI, *Professor of Law.* A.B., Bowdoin College; J.D., University of Chicago.
HELEN P. GARFIELD, *Professor of Law.* B.S.J., Northwestern University; J.D., University of Colorado.
HAROLD GREENBERG, *Professor of Law.* A.B., Temple University; J.D., University of Pennsylvania.
JEFFREY W. GROVE, *Professor of Law.* A.B., Juniata College; J.D., George Washington University.
FRANCES W. HARDY, *Clinical Assistant Professor of Law.* B.S., Ball State University; J.D., Indiana University-Indianapolis.
WILLIAM F. HARVEY, *Carl M. Gray Professor of Law & Advocacy.* A.B., University of Missouri; J.D., LL.M., Georgetown University.
PAUL T. HAYDEN, *Assistant Professor of Law.* B.A., Yale University; J.D., University of California, Los Angeles.
W. WILLIAM HODES, *Professor of Law.* A.B., Harvard College; J.D., Rutgers Newark.
LISA C. IKEMOTO, *Assistant Professor of Law.* B.A., University of California, Los Angeles; J.D., University of California, Davis; LL.M., Columbia University.
LAWRENCE A. JEGEN, III, *Thomas F. Sheehan Professor of Tax Law and Policy.* A.B., Beloit College; J.D., M.B.A., The University of Michigan; LL.M., Harvard University.
HENRY C. KARLSON, *Professor of Law.* A.B., J.D., LL.M., University of Illinois.
WILLIAM ANDREW KERR, *Professor of Law.* A.B., West Virginia University; J.D., LL.M., Harvard University; D.B., Duke University.

ELEANOR D. KINNEY, *Professor of Law and Director of the Center for Law and Health.* B.A., Duke University; M.A., University of Chicago; J.D., Duke University; M.P.H., University of North Carolina.

WALTER W. KRIEGER, *Associate Professor of Law.* A.B., Bellarmine College; J.D., University of Louisville; LL.M., George Washington University.

NORMAN LEFSTEIN, *Dean and Professor of Law.* LL.B., University of Illinois; LL.M., Georgetown University.

DEBORAH McGREGOR, *Lecturer in Law.* B.A., University of Evansville; J.D., Georgetown University.

WILLIAM E. MARSH, *Professor of Law.* B.S., J.D., University of Nebraska.

SUSANAH M. MEAD, *Professor of Law.* B.A., Smith College; J.D., Indiana University-Indianapolis.

MIRA MIHAJLOVICH, *Lecturer in Law.* B.A., Indiana University; M.P.A., University Texas; J.D., Indiana University-Bloomington.

MARY H. MITCHELL, *Professor of Law.* A.B., Butler University; J.D., Cornell Law School.

JAMES P. NEHF, *Assistant Professor of Law.* B.A., Knox College; J.D., University of North Carolina.

THOMAS R. NEWBY, *Lecturer in Law.* A.B., J.D., Indiana University-Bloomington.

KIMBERLY E. O'LEARY, *Clinical Assistant Professor of Law.* B.A., Oberlin College; J.D., Northwestern University.

DAVID R. PAPKE, *Professor of Law and American Studies.* A.B., Harvard College; J.D., Yale University; M.A. in American Studies, Yale University; Ph.D. in American Studies, The University of Michigan.*

RONALD W. POLSTON, *Professor of Law.* B.S., Eastern Illinois University; LL.B., University of Illinois.

JOAN M. RUTHENBERG, *Director of Legal Writing.* B.A., Mississippi University for Women; J.D., Indiana University-Indianapolis.

ANDREW T. SOLOMAN, *Lecturer in Law.* A.B., University of Michigan; J.D., Boston University.

DAVID E. SORKIN, *Lecturer in Law.* B.S. Bus., B.A., Indiana University; J.D., Harvard Law School.

KENNETH M. STROUD, *Professor of Law.* A.B., J.D., Indiana University-Bloomington.

JAMES W. TORKE, *Professor of Law.* B.S., J.D., University of Wisconsin.

JOANNE ORR VAN PELT, *Visiting Assistant Professor.* J.D., California Western.

JAMES PATRICK WHITE, *Professor of Law and Consultant on Legal Education to the American Bar Association.* A.B., University of Iowa; J.D., LL.M., George Washington University.

LAWRENCE P. WILKINS, *Professor of Law.* B.A., The Ohio State University; J.D., Capital University Law School; LL.M., University of Texas.

MARY T. WOLF, *Director of Clinical Programs.* B.A., Saint Xavier College; J.D., University of Iowa College of Law.

Emeriti

AGNES P. BARRETT, *Associate Professor Emerita.* B.S., J.D., Indiana University.

CLEON H. FOUST, *Professor Emeritus.* A.B., Wabash College; J.D., University of Arizona.

JOHN S. GRIMES, *Professor of Jurisprudence Emeritus.* A.B., J.D., Indiana University.

MELVIN C. POLAND, *Cleon H. Foust Professor of Law Emeritus.* B.S., Kansas State University; LL.B., Washburn University; LL.M., University of Michigan.

R. BRUCE TOWNSEND, *Cleon H. Foust Professor of Law Emeritus.* A.B., Coe College; J.D., University of Iowa.

Adjunct Faculty

CYNTHIA M. ADAMS, J.D., Indiana University-Indianapolis.
RICHARD P. GOOD, J.D., Indiana University-Indianapolis.
RALPH F. HALL, J.D., University of Michigan.
GRANT W. HAWKINS, J.D., Indiana University-Indianapolis.
THOMAS Q. HENRY, J.D., Indiana University-Indianapolis.
SWADESH S. KALSI, LL.M., George Washington University.
LARRY A. LANDIS, J.D., Indiana University-Indianapolis.
JANE MAYNUS, J.D., Indiana University-Indianapolis.
SHARON F. MURPHY, J.D., Indiana University-Indianapolis.
LAURA C. NEHF, J.D., University of Georgia.
GARY P. PRICE, J.D., Indiana University-Indianapolis.
THOMAS R. RUGE, J.D., Valparaiso University.
GEOFFREY SEGAR, J.D., Indiana University-Bloomington.
PATRICIA P. WAGNER, J.D., Indiana University-Indianapolis.

Law Library Faculty

JAMES F. BAILEY III, *Professor and Director of Law Library*. A.B., J.D., M.A.L.S., University of Michigan.
MINDE C. GLENN, *Assistant Director for Readers' Services*. B.A., Western Michigan University; M.L.S., Indiana University.
WENDELL E. JOHNTING, *Catalog Librarian*. A.B., Taylor University; M.L.S., Indiana University.
MAHNNAZ K. MOSHFEGH, *Acquisition/Serials Librarian*. B.A., National University of Iran; M.S., Tehran University; M.A., Ball State University; M.L.S., Ph.D., Indiana University.
KIYOSHI OTSU, *Computer System Specialist and Catalog Librarian*. A.A., Parkland College; A.B., M.S., C.A.S., University of Illinois.**
TERENCE L. ROSE, *Reference Librarian*. B.G.S., M.L.S., University of Michigan.

* On leave 1991-92.

** On leave at Tenri University, Nara, Japan, 1991-92.

Indiana Law Review

Volume 25

1991

Number 1

ARTICLES

Transfers Prior to Marriage and the Uniform Probate Code's Redesigned Elective Share—Why the Partnership is Not Yet Complete

RENA C. SEPLOWITZ*

INTRODUCTION

Elective share statutes have recently attracted attention from state legislatures, law revision commissions, model code drafters, and commentators.¹ Most notable is the 1990 revision of Article II of the

* Associate Professor of Law, Touro College, Jacob D. Fuchsberg Law Center. A.B., 1973, Barnard College; J.D., 1975, Columbia University. The author gratefully acknowledges the valuable comments of Professors Dan Subotnik and Sidney Kwestel and the contribution of her research assistants, David Desmond, Michael Wasserman, Lainie Dolinger, Murray Bach, Diedre Yablon, Michael Beer, and Denise Saggio.

1. See, e.g., S.C. CODE ANN. §§ 62-2-201, -202, -205 (Law Co-op. Supp. 1990) (the Reporter's Comments to the 1987 version of § 62-2-202 note that the section "rejects the 'augmented estate' concept promulgated by the drafters of the Uniform Probate Code as unnecessarily complex," although this approach was recommended in LeBlanc, *The Proposed South Carolina Probate Code*, 36 S.C.L. REV. 511 (1985)); 20 PA. CONS. STAT. ANN. § 2203 (Purdon Supp. 1991); WIS. STAT. ANN. § 861.17 (West 1991); UNIF. PROB. CODE art. II, pt. 2, 8 U.L.A. 78 (Supp. 1991) [hereinafter U.P.C.]; FIRST REPORT OF THE NEW YORK ESTATES POWERS AND TRUSTS LAW ADVISORY COMMITTEE (1991) [hereinafter ADVISORY COMMITTEE REPORT]; NEW YORK LEGIS. COMM'N REPORTS AND NEW YORK STATE BAR ASSOCIATION REPORT ON NEW YORK EST. POWERS AND TRUSTS LAW § 5-1.1 (1990) [hereinafter N.Y. BAR REPORT]; Kwestel & Seplowitz, *Testamentary Substitutes: Retained Interests, Custodial Accounts and Contractual Transactions — A New Approach*, 38 AM. U.L. REV. 1 (1988); Langbein & Waggoner, *Redesigning the Spouse's Forced Share*, 22 REAL PROP. PROB. & TRUST J. 303 (1987); Oldham, *Should the Surviving Spouse's Forced Share Be Retained?*, 38 CASE W. RES. L. REV. 223 (1987); Waggoner, *The Multiple-Marriage Society and Spousal Rights Under the Revised Uniform Probate Code*, 76 IOWA L. REV. 223 (1991); Waggoner, *Spousal Probate Rights in a Multiple-Marriage Society*, 45 REC. A.B. CITY N.Y. 339, 340 & n.6 (1990) [hereinafter Waggoner, *Spousal Probate Rights*]; Note, *Uniform Probate Code Section 2-202: A Proposal to Include Life Insurance Assets Within the Augmented Estate*, 74 CORNELL L. REV. 511 (1989).

Uniform Probate Code (UPC)² which embodies a profound philosophical change in the concept and structure of the elective share in common-law jurisdictions. Virtually alone among elective share statutes, the UPC grapples with the elusive nature of marriage and takes into account equitable distribution and community property models.³ No longer are the rationales of the elective share system, marriage as an economic partnership, and support of the surviving spouse,⁴ divorced from the specific provisions of a forced share statute. In particular, the partnership, or marital-sharing theory of marriage, which is increasingly the primary justification for the elective share (as well as for equitable distribution and community property), underlies the redesign of the UPC.⁵

In a revolutionary move for an elective share statute, the UPC combines the assets of both spouses in computing the augmented estate⁶ against which the survivor's right of election is measured.⁷ To achieve a more equitable result, particularly with respect to late-in-life multiple marriages, the UPC bases the surviving spouse's fractional share on the length of the marriage.⁸ These features are designed to recognize

2. U.P.C. art. II, pt. 2 (Supp. 1991) (revised Article II was drafted by the National Conference of Commissioners on Uniform State Laws and was approved in July, 1990).

3. See Wis. STAT. ANN. §§ 861.01 to -.13 (West 1991) (adopting a modified form of the Uniform Marital Property Act); Oldham, *supra* note 1, at 245-46 (recommending basing the elective share on a concept of marital partnership that looks to divorce law to determine marital property or replacing the elective share with the Uniform Marital Property Act); Comment, *Spousal Disinheritance: The New York Solution — A Critique of Forced Share Legislation*, 7 W. NEW ENG. L. REV. 881, 905-07 (1985) (arguing for equitable distribution in cases of disinheritance); Apfel, *Divorce and Death: Disparity in Economic Rights of Spouse*, N.Y.L.J., Jan. 28, 1988, at 1, col. 3 (advocating use of equitable distribution as a model for forced share legislation).

4. See Carr v. Carr, 120 N.J. 336, 349-50, 576 A.2d 872, 879 (1990) (the principle behind both the probate and equitable distribution statutes is that each spouse's efforts during marriage contribute to and create marital assets); Bialczak v. Moniak, 373 Pa. Super. 251, 255, 540 A.2d 962, 965 (1988) (Pennsylvania legislature "intended to establish a comprehensive statutory system of providing a measure of economic security for surviving spouses"); U.P.C. art. II, pt. 2 general comment (Supp. 1991); Kwestel & Seplowitz, *supra* note 1, at 3 & n.5.

5. See U.P.C. art. II, pt. 2 general comment (Supp. 1991); Waggoner, *Spousal Probate Rights*, *supra* note 1, at 349. The 1990 UPC is also referred to as the "redesigned" UPC.

6. U.P.C. § 2-202 (Supp. 1991).

7. Waggoner, *Spousal Probate Rights*, *supra* note 1, at 361.

8. U.P.C. § 2-201(a) (Supp. 1991); Waggoner, *Spousal Probate Rights*, *supra* note 1, at 360. See also Langbein & Waggoner, *supra* note 1, at 314-21 (advocating that the pre-1990 UPC be modified to base the elective share on the length of the marriage and the amount of property owned by the surviving spouse). The UPC avoids the difficult administrative problems of tracing assets by applying a percentage based on the length of the marriage to the spouse's combined assets. For a discussion of the UPC's accrual system, see *infra* notes 194-96 and accompanying text.

each spouse's pecuniary and nonpecuniary contributions⁹ to the marriage, the expectation that marital assets will be shared regardless of the name in which they are titled, and lost opportunities.¹⁰ The pre-1990 UPC focused on an elective share that was based on the decedent's estate augmented by certain inter vivos transfers received by the surviving spouse from the decedent.¹¹ By combining the estates of both spouses, the redesigned UPC rejects the traditional elective share scheme and incorporates significant features of community property systems. In so doing, it seeks to avoid some of the inequities inherent in the forced share in many common-law jurisdictions and to minimize disturbances of the decedent's estate plan. Because a traditional elective share statute does not recognize that property frequently is titled arbitrarily in one spouse's name, the survivor may be undercompensated or overcompensated.¹² The UPC, in effect, casts off the shackles of dower and the forced share that generally give the surviving spouse a one-third interest in some or all of the decedent's property regardless of the extent of the survivor's own estate and the proportion it bears to the decedent's estate.

In addition to implementing the policy of marital-sharing, redesigned Article II, even more than its predecessor, is concerned with preventing circumvention of the surviving spouse's right of election.¹³ The inclusion in the reclaimable estate¹⁴ of certain property arrangements that were formerly excluded from the right of election, principally

9. Nonpecuniary contributions include care of the children and the house.

10. See U.P.C. art. II, pt. 2 general comment (Supp. 1991); Waggoner, *Spousal Probate Rights*, *supra* note 1, at 349-50.

11. The pre-1990 UPC included probate assets and testamentary substitutes such as revocable trusts and joint tenancies with right of survivorship in the decedent's estate. U.P.C. § 2-202(1) (1983).

12. Undercompensation occurs if most of the marital assets are titled in the decedent's name. Overcompensation occurs if the majority of the assets are titled in the surviving spouse's name. Overcompensation may also occur if the surviving spouse is the beneficiary of will substitutes that cannot be included in the elective share. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(2) (McKinney 1981) (excluding insurance proceeds even if payable to the surviving spouse); 20 PA. CONS. STAT. ANN. § 2203(b)(2) (Purdon Supp. 1991) (excluding insurance proceeds even if payable to the surviving spouse). The Advisory Committee Report recommended amending § 5-1.1 to include insurance proceeds in the net estate if the decedent retained the incidents of ownership. ADVISORY COMMITTEE REPORT, *supra* note 1, at 15.

13. Compare U.P.C. § 2-202 & comment (Supp. 1991) with U.P.C. § 2-202(1) (1983).

14. U.P.C. § 2-202(b)(2) defines the reclaimable estate as the value of certain inter vivos property arrangements made or held by the decedent. For a discussion of testamentary substitutes includible regardless of when created, see *infra* notes 203-13 and accompanying text.

insurance proceeds payable to persons other than the surviving spouse and property subject to a presently exercisable general power of appointment, represents a significant step in limiting the types of evasive devices a decedent may use to disinherit a spouse.¹⁵

For the most part, the modifications wrought by the UPC are positive. Nevertheless, problems remain. In its embrace of the concept of marriage as an economic partnership, the UPC raises, but does not always respond to, several critical issues. Some of these questions relate to the nature of the marital partnership and the definition of partnership assets. Others involve the classification of property acquired prior to the formation of the partnership or marriage (although not technically made part of the partnership's assets by being titled in both names or explicitly used for maintenance and support of the partnership unit, *i.e.*, the family). Resolution of these issues requires an analysis of: (1) whether spousal rights after dissolution of the marriage by divorce or by death should be treated differently; (2) whether the notion of separate property¹⁶ should be rejected in defining the scope of the right of election, a result that would promote the concept of marriage as a total partnership;¹⁷ and (3) whether the legitimate interests of children from a prior marriage¹⁸ can be accommodated in a total partnership model. The UPC, in accord with the predominant equitable distribution and community property models, appears to endorse the view that marriage is a limited partnership with separate property theoretically excluded from the partnership.

Another problem area in the revised UPC is the continued opportunity for spousal disinheritance as a result of statutory omissions. For example, a decedent, during life, may enjoy the benefits of property either by consumption or by economic control over its disposition.

15. See U.P.C. § 2-202(b)(2) (Supp. 1991) (including insurance proceeds, annuities, and other transfers in the augmented estate); Kwestel & Seplowitz, *supra* note 1 (analyzing the types of retained interest transfers, contractual arrangements, and custodial accounts that should be subject to the right of election); Note, *supra* note 1 (advocating the includability of insurance proceeds in the elective share).

16. The concept of separate property, embodied in community property and equitable distribution systems, generally refers to property acquired prior to marriage or during marriage by gift or inheritance.

17. The view of marriage as a complete partnership in which all assets of both spouses are combined at termination, known as the "deferred community," has been adopted in the Nordic countries, Israel, Quebec, Germany, and Switzerland. See M. GLENDON, THE TRANSFORMATION OF FAMILY LAW 131-34 (1989).

18. The UPC is concerned with preserving the interests of children from a prior marriage in property acquired during that marriage. See U.P.C. art. II, pt. 2 general comment (Supp. 1991); Waggoner, *Spousal Probate Rights*, *supra* note 1, at 354. A decedent may also have an interest in not allowing a subsequent spouse to disturb property arrangements for the benefit of elderly parents.

Under both versions of the Code, testamentary substitutes¹⁹ are generally only subject to the right of election if the decedent has a beneficial interest in the property.²⁰ This allows a decedent to create a trust and reserve a power to invade the corpus for the benefit of a person other than the spouse. The decedent has retained economic control over the principal of the trust and sheltered unneeded assets from the claims of the surviving spouse. To the extent that the UPC characterizes the principal as marital partnership assets, that is, property acquired during marriage, this result does not promote the stated goals of the redesigned elective share.

Contrary to the marital-sharing theory, this framework enables a decedent to prevent the surviving spouse from sharing in partnership property at the decedent's death and to remove assets necessary for the survivor's support. In addition, a statute that requires the retention of a beneficial interest disregards the idea that ownership of property not only means having the power to consume or enjoy its benefits, but also connotes an ability to determine the beneficiaries of the property and the timing and amounts of its distribution.²¹ Thus, the ability to dispose of property should be as significant a characteristic for purposes of the elective share as it is for federal estate tax purposes.²²

In addition, the UPC largely ignores transfers prior to marriage, whether outright or in the form of testamentary substitutes.²³ Insulation

19. In this Article, the term "testamentary substitute" refers to an *inter vivos* transfer made without full and adequate consideration that may be subject to the surviving spouse's right of election.

20. U.P.C. § 2-202(b)(2)(iv) (Supp. 1991); U.P.C. § 2-202(1) (1983). Under the pre-1990 Code, the decedent must have retained a beneficial interest in the principal. The 1990 version includes beneficial interests in income or principal. U.P.C. § 2-202(b)(2)(iv)(B) (Supp. 1991). Under the pre-1990 UPC, however, testamentary substitutes were includible in the augmented estate only if they were created during marriage. U.P.C. § 2-202(1)(ii) (1983). This continues to be true under the new Code except for certain insurance policies, unilaterally severable jointly held property with right of survivorship, and property subject to a general presently exercisable power of appointment. These property arrangements are includible regardless of when they were created. U.P.C. § 2-202(b)(2) (Supp. 1991).

21. This is the premise underlying the recommendations for statutory change in Kwestel & Seplowitz, *supra* note 1, at 57-66.

22. The ability to dispose of and enjoy property provides the basis for including certain transfers in the gross estate for federal estate tax purposes. See I.R.C. §§ 2036-38, 2040-42 (1986). For an elective share statute based on the federal estate tax approach, see DEL. CODE ANN. tit. 12, §§ 901-02 (1987 & Supp. 1990). See also Kwestel & Seplowitz, *supra* note 1, at 6 n.17 (discussing authorities that suggest using federal estate tax law as a basis for forced share statutes); Waggoner, *Spousal Probate Rights*, *supra* note 1, at 378 n.68.

23. Wisconsin and Missouri, which codified the common-law doctrine of transfers in fraud of marital rights, are the exceptions to the approach used by other elective share

of these transfers, particularly those that are will substitutes, creates a significant opportunity for abuse. For example, prior to marriage and possibly in contemplation of marriage, a decedent may create a revocable trust or an irrevocable trust, reserving a life estate and a power to invade the principal for the decedent's support and maintenance. The surviving spouse will be unable to elect against either of them. Both trusts, however, will be subject to the surviving spouse's right of election if created during marriage. When the decedent's property interests in the trusts are the same during marriage, the trust property, to the extent of the decedent's power, constitutes assets of the marital partnership. Fairness mandates identical treatment of the trusts regardless of whether the spouse was a first or subsequent spouse.

Sheltering these dispositions from the elective share may encourage some individuals to create will substitutes prior to marriage to the detriment of their spouses without serious detriment to themselves. The UPC's continued failure to deal with this possibility frustrates an avowed purpose of both versions (and any other progressive elective share statute) to "protect the surviving spouse against so-called 'fraud on the spouse's share.'"²⁴ In fact, this treatment of premarriage will substitutes also ignores an entire body of law that gives a spouse an interest in many of these transfers and that could have served as a model for the redesigned UPC. Under the doctrine of antenuptial transfers in fraud of marital rights, certain premarital conveyances, both outright and in the form of will substitutes, have been invalidated in whole or in part to prevent spousal disinheritance. The consequences of excluding these transfers and ignoring this doctrine are illustrated by the following three hypotheticals that serve as focal points throughout this Article.

Case 1

Horace had a first wife, Fran, who died when he was sixty years old. During their thirty-five year marriage, they held most of their assets jointly. After Fran died, Horace made several substantial dispositions of his property. His main purpose in making these transfers was to provide for himself for the remainder of his life and for his children after death without having most of his assets pass through his probate estate. Horace transferred \$350,000 to an irrevocable inter vivos trust pursuant to which he was to receive the income for life and such sums from the principal as he and his son, Sam, deemed

statutes, including the UPC. *See* Mo. ANN. STAT. § 474.150(1) (Vernon Supp. 1990); Wis. STAT. ANN. § 861.17 (West 1991). In addition, the redesigned UPC includes certain inter vivos property arrangements in the elective share regardless of when they were created. U.P.C. § 2-202(b)(2)(i)-(iii) (Supp. 1991).

24. U.P.C. § 2-202 comment (Supp. 1991).

necessary for his support and maintenance. Upon his death, the remaining principal was to be paid to his daughter, Doris, if living, or if not, to her surviving issue. He transferred his house, previously owned with Fran as tenants by the entirety and appraised at her death at \$350,000, to his son, Sam, reserving a life estate for himself.

Three years after Horace made these transfers, he married Wanda. Up to the time of his marriage, Horace periodically transferred money to the trust. Horace and Wanda lived in his house during their marriage. Wanda did not know that Horace owned only a life estate. She was also unaware of the trust's existence. At Horace's death fifteen years later, he was survived by Wanda and his two children. He left a probate estate of \$20,000 and assets held jointly with Wanda with right of survivorship of \$100,000. The real property and trust principal were worth \$900,000. In his will, Horace left one-half of his estate to Wanda, who had approximately \$50,000 of her own assets.

Case 2

Consider the case of Wilma who was sixty years old at the death of her first husband, Felix. They were married for thirty years and had two children. Felix left all of his property to Wilma in his will. After his death, Wilma kept most of her assets (worth approximately \$800,000) in her own name. At age sixty-five, Wilma married Hubert and subsequently executed a will leaving one-half of her estate to him and the other half to her surviving issue. She frequently stated that she would like her children to be well provided for after her death. Following a fifteen year marriage, Wilma died, leaving Hubert, her two children, and a probate estate valued at \$1,000,000. At Wilma's death, Hubert had approximately \$50,000 titled in his name and no property in the form of will substitutes.

Case 3

Winnifred and Fred were married for thirty-three years and had two children. When Winnifred was sixty-two years old, Fred died, leaving her \$400,000 in cash and securities and a house worth \$350,000. At Fred's death, Winnifred had approximately \$200,000 in her own name. Five years later, Winnifred married Horatio. After their marriage, Winnifred transferred the house to her daughter, Dorothy, reserving a life estate. She also created an irrevocable inter vivos trust with a corpus of \$400,000, naming herself and Dorothy as trustees. Pursuant to the trust, the trustees were to pay Winnifred the income for life and at her death, the principal to her son, Steven. During Winnifred's life, the trustees had the power to make discretionary principal payments to Winnifred and Steven. Winnifred's motives in making these transfers were to provide for her children and to reduce the size of her probate estate. Horatio and Winnifred were married for fifteen years when Winnifred died. She left her entire probate estate of \$50,000 to Horatio who had approximately \$50,000 of his own assets.

Despite the fact that in each of these three cases the decedent's estate planning goal was to provide for children of a prior marriage and not to disinherit the subsequent spouse, the UPC does not treat these cases consistently. Their treatment is different because the UPC includes certain will substitutes in the elective share only if they are created during marriage. In the first case, under the UPC's elective share provisions, Wanda would receive virtually nothing as a consequence of Horace's death.²⁵ Horace had a small probate estate, a modest amount of property held jointly with Wanda with a right of survivorship, and a large amount of property in the form of will substitutes created prior to marriage. In the second case, however, Hubert was comfortably provided for because Wilma did not create any will substitutes. Even if Wilma left Hubert nothing in her will, he would receive approximately fifty percent of her large probate estate by exercising his right of election. In Case 3, the UPC grants Horatio a right of election because of the testamentary substitutes established by Winnifred during their marriage.

These results appear contradictory because Horace, Wilma, and Winnifred had essentially the same power over and interest in the property during life and at death. In each case, the decedent's economic power over or use of the property was equivalent to ownership.²⁶ Furthermore, their family situations were comparable: Horace, Wilma, and Winnifred were primarily concerned with providing for the children of their first marriage and did not have the express intention of disinheriting his or her surviving spouse. Thus, there seems to be no reason for an elective share statute to distinguish among the three cases.

In its emphasis on the interests of the children from a prior marriage, the UPC overlooks the contribution of the late-in-life surviving spouse and the fact that the property used to support the older couple is income frequently derived from premarriage property. The rights of the beneficiaries of the transferred property, often children from a prior marriage, can still be protected under an approach that balances the interests of the surviving spouse and the children.²⁷ As with testamentary substitutes created during marriage, their interest will be subject to diminution, not elimination, by the exercise of a right

25. This would not be true if the will substitutes were created during marriage as in Case 3. See N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1) (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203 (Purdon Supp. 1991); U.P.C. § 2-202(b)(2)(iv) (Supp. 1991).

26. See *supra* notes 21-22 and accompanying text.

27. But cf. Wis. STAT. ANN. § 861.17(2) (repealed 1985) (providing that "fraudulent" premarriage transfers for the benefit of children of a former marriage are not subject to the right of election). For the text of this section, see *infra* note 129.

of election. Furthermore, the children's interests in premarriage (or post-wedding) transfers can be secured if, after disclosure by Horace (or Winnifred), Wanda (or Horatio) consents to them and waives the right of election.²⁸ The UPC is not alone in this inconsistent treatment of testamentary substitutes. Many elective share statutes either ignore testamentary substitutes entirely or include only a limited number in the augmented estate.

Case 1 involves transfers prior to the contemplation of marriage or engagement. Assume for a moment that in Case 1, Horace created the will substitutes after his engagement, but before his marriage to Wanda. Would the proximity of the transfers to the upcoming marriage produce a different result under the UPC? Are Horace's additions to the irrevocable trust during the engagement relevant? Nothing in the UPC or its comments suggests that transfers in contemplation of marriage are subject to the right of election. Nevertheless, under the common-law doctrine of antenuptial transfers in fraud of marital rights, these transfers could be invalidated under the appropriate circumstances. For example, if Horace made these transfers for the express purpose of defrauding Wanda of her right to an elective share, the transfers are voidable. Although the doctrine is antithetical to the UPC's objective approach to testamentary substitutes, it deals with an important area ignored by the UPC and most elective share statutes.

A principal premise of this Article is that any elective share system designed to prevent or to deter spousal disinheritance must critically examine the appropriate treatment of antenuptial conveyances. Its endorsement of the UPC's inclusion of testamentary substitutes based on objective criteria precludes support for a doctrine grounded on a decedent's subjective intent. Because an elective share statute should focus on the decedent's power over and interest in property at death, the time of the creation of a testamentary substitute should be immaterial.²⁹ Regardless of whether Horace created the trust and life estate either before he met Wanda or on the eve of marriage, this author takes the position that the UPC (or any other elective share statute) should give Wanda a right to elect against these transfers. Due to the special nature of the life estate, subjecting the principal to the right of election may not be desirable when a decedent created the life estate prior to marriage. Instead, consideration should be given to granting the sur-

28. See U.P.C. § 2-204 (Supp. 1991) (allowing written waiver of right of election). Disclosure of antenuptial transfers without the prospective spouse's written consent is arguably sufficient to bar a right of election. Requiring a written waiver would seem preferable because it should produce less litigation and a clearer indication of the surviving spouse's understanding.

29. But see U.P.C. § 2-202 comment (1983).

viving spouse a right to elect against the income interest for the remainder of the spouse's life.³⁰

This Article begins with an examination of the historical background of the common-law doctrine of antenuptial transfers in fraud of marital rights and its relationship to the elective share. Although this Article rejects the specific approach of this doctrine, the author finds that it provides invaluable insights into both intentional and unintentional spousal disinheritance and serves as an important guide in correcting a serious omission in the UPC and other elective share statutes. In the majority of cases invalidating an antenuptial conveyance, the decedent retained an interest in or power over the transferred property, although this element was never a formal requirement of the doctrine. Surprisingly, many courts are often more sensitive to the retention of interests and powers in premarriage than in marital transfers. Next, the Article examines several statutory approaches, including certain prenuptial transfers in the forced share, that may serve as alternative models to the common-law doctrine.

The Article will then analyze and evaluate the UPC's adoption of the marital-sharing theory of marriage, discuss a total partnership theory as an alternative, and examine the UPC's treatment of prenuptial transfers of property. Specifically, it will focus on whether the exclusion of these prenuptial transfers promotes the goals of marital partnership or support. In the course of its examination of the common-law doctrine and the statutory approaches, the Article will analyze the deficiencies of each model in light of the policies of the elective share. Finally, it will propose guidelines that better accomplish these goals and close a significant loophole in the UPC and other statutory and common-law schemes. The guidelines used should effectively and equitably protect the surviving spouse, produce greater certainty in property arrangements for the benefit of the decedent's family, and reduce the likelihood of litigation.

I. THE PROBLEM PRESENTED—TOWARD AN INTEGRATED TREATMENT OF TRANSFERS PRIOR TO AND DURING MARRIAGE

Transfers of property by prospective spouses to third persons have long been a concern of legal systems.³¹ Originally, a woman was

30. For a discussion of three approaches to the includibility of retained life estates created prior to marriage, see *infra* pages 60-61.

31. For example, under Jewish law, a woman may generally dispose of her property before marriage. If she misrepresents what she has done to her intended husband, however, her fraud may provide a basis for annulling the marriage on the theory that she entered the marriage with a concealed defect. The transfer itself, on the other hand, is valid. See

motivated to make such transfers because her husband acquired the right to manage and consume her property upon marriage.³² For the prospective husband and wife, the transfer was generally designed to prevent the spouse from realizing any interest in his or her property upon death. If the transfer in either case was a fraud on the spouse's marital rights, regardless of the decedent's retention of any interest in or control over the property, it could be invalidated in whole or in part under the doctrine of antenuptial transfers in fraud of marital rights.³³ This doctrine, which originated in England to protect husbands from transfers of property made by their prospective wives,³⁴ was applied in the United States³⁵ largely against husbands who transferred

BABYLONIAN TALMUD, SEDER NASHIM II: TRACTATE KETHUBOTH 78a (Soncino ed. 1936). As a general rule, misrepresentation of one's financial condition does not furnish a ground for annulment in this country. *See, e.g.*, Woronzoff-Daschkoff v. Woronzoff-Daschkoff, 303 N.Y. 506, 104 N.E.2d 877 (1952); Francis v. Francis, 21 V.I. 263, 264-65, 11 F.L.R. 1269 (1985). To annul a marriage based on fraud, courts generally require misrepresentation with respect to: (1) pregnancy (Reynolds v. Reynolds, 3 Allen 605 (Mass. 1862)); (2) physical or mental health (Stone v. Stone, 136 F.2d 761 (D.C. Cir. 1943)); or (3) intent to consummate the marriage (Marshall v. Marshall, 212 Cal. 736, 300 P. 816 (1931); Security-First Nat'l Bank v. Schaub, 71 Cal. App. 2d 467, 162 P.2d 966 (1945); Chipman v. Johnston, 237 Mass. 502, 130 N.E. 65 (1921); Anders v. Anders, 224 Mass. 438, 113 N.E. 203 (1916); Schonfeld v. Schonfeld, 260 N.Y. 477, 184 N.E. 60 (1933); Avnery v. Avnery, 50 A.D.2d 806, 375 N.Y.S.2d 888 (1975)). *But see* Comment, *Husband and Wife—Conveyances Before Marriage as Fraud on Wife*, 8 ILL. L. REV. 500 (1913) (suggesting that annulment is the proper remedy for an antenuptial transfer in fraud of marital rights).

32. *See infra* text accompanying notes 73-75.

33. For a discussion of this doctrine, see W. MACDONALD, *FRAUD ON THE WIDOW'S SHARE* 354-65 (1960); Bregy & Wilkinson, *Antenuptial Transfers as Frauds on Marital Rights in Pennsylvania*, 90 U. PA. L. REV. 62 (1941); Lowe, *Transfers in Fraud of Marital Rights*, 26 Mo. L. REV. 1 (1960) (discussing the doctrine's application in Missouri); Vaughan, *Antenuptial Conveyance or Agreement Intended to Defeat Right of Election*, 32 ST. JOHN'S L. REV. 174 (1958) (discussing the doctrine's application in New York).

34. *See, e.g.*, St. George v. Wake, 1 My. & K. 610, 39 Eng. Rep. 812 (1833); Goddard v. Snow, 1 Russ. 485, 38 Eng. Rep. 187 (1826); DeManneville v. Crompton, 1 V. & B. 354, 35 Eng. Rep. 138 (1813); Strathmore v. Bowes, 1 Ves. jun. 22, 30 Eng. Rep. 211 (1789), *aff'd sub nom.* Bowes v. Bowes, 6 Brown 427, 2 Eng. Rep. 1178 (1797); King v. Cotton, 2 P. Wms. 674, 24 Eng. Rep. 912 (1732); Carlton v. Earl of Dorset, 2 Vern. 17, 23 Eng. Rep. 622 (1686); Hunt v. Matthews, 1 Vern. 408, 23 Eng. Rep. 549 (1686); Howard v. Hooker, 2 Chan. Rep. 81 (1672-73).

35. Historically, in the United States, the doctrine was applied to invalidate transfers by both prospective husbands and wives. *See, e.g.*, Swaine v. Perine, 5 Johns. Ch. 482 (N.Y. 1821) (prospective husband's transfer to daughter of a deed fraudulent as to wife's dower rights and as to subsequent creditors); Littleton v. Littleton, 18 N.C. 327, 330 (1835) (invalidating deed that was made 15 days before marriage for the benefit of the children of a prior marriage to the extent of the widow's dower interest based on a statute providing that "conveyances made fraudulently . . . with the intention to defeat the widow of her dower . . . shall be held and deemed void"). At least one case, however, indicated

their real property,³⁶ and sometimes their personality, prior to marriage.³⁷ Thus, although the doctrine was initially used in the United States to safeguard the husband's rights in his wife's property,³⁸ it was applied in most instances to preserve the wife's inchoate right of dower.³⁹ Subsequently, the doctrine was expanded in a number of jurisdictions to protect the surviving spouse's forced share in the decedent's real and personal property.⁴⁰

The elements of fraud required by the doctrine differed markedly from those comprising the usual fraud case.⁴¹ Because the rights protected by the common-law doctrine did not arise until marriage or death, the prospective spouse did not have any existing property interests of which he or she could be defrauded.⁴² In effect, any misrepresentation or non-

in dicta that the doctrine only applied to prenuptial transfers by women. *See Peay v. Peay*, 2 S.C. 409 (1844). This statement was harshly criticized in *Brooks v. McMeekin*, 37 S.C. 285, 304-05 (1892), on the ground of safeguarding a woman's right to be supported by her husband.

36. *See, e.g.*, *Chandler v. Hollingsworth*, 3 Del. Ch. 99 (1867).

37. *See, e.g.*, *LeStrange v. LeStrange*, 242 A.D. 74, 273 N.Y.S. 21 (1934).

38. *See, e.g.*, *Poston v. Gillespie*, 58 N.C. 258 (1859); *Ramsay v. Joyce*, 1 McMul. Eq. 236 (S.C. 1841).

39. *See, e.g.*, *Chandler*, 3 Del. Ch. at 99 (applying the doctrine only to antenuptial transfers of realty).

40. *See, e.g.*, *LeStrange*, 242 A.D. at 74, 273 N.Y.S. at 21.

41. Actual fraud has been defined as an "intentional deception to induce another to part with property or to surrender some legal right, and which accomplishes the end designed." *Stanley v. Sewell Coal Co.*, 169 W. Va. 72, 76, 285 S.E.2d 679, 683 (1981). Actual fraud differs from constructive fraud which does not require proof of a fraudulent intent to protect important societal interests. *Id. See also Perlberg v. Perlberg*, 18 Ohio St. 2d 55, 247 N.E.2d 306 (1969). Many courts that applied the doctrine of transfers in fraud of marital rights to invalidate a prenuptial conveyance did so on a finding of constructive fraud. *See, e.g.*, *Chandler v. Hollingsworth*, 3 Del. Ch. 99, 113 (1867) (in the absence of misrepresentation, nondisclosure of an antenuptial conveyance will support a finding of constructive fraud and will invalidate the portion of the transfer that is necessary to protect the wife's dower rights); *Cranson v. Cranson*, 4 Mich. 230 (1856) (invalidating a prospective husband's transfer of realty to his children prior to marriage on the ground that he did not deliver the deed until after the marriage, but even if he had delivered it before marriage, it would have been fraudulent because it was executed secretly for the purpose of cutting off her dower); *Arnegaard v. Arnegaard*, 7 N.D. 475, 75 N.W. 797 (1898) (invalidating a husband's transfer of real estate with a reserved life estate two months before marriage to the extent of the wife's homestead right on the ground that the husband failed to disclose the transfer); *Ward v. Ward*, 63 Ohio St. 125, 57 N.E. 1095 (1900) (invalidating a prospective husband's transfers of land to three of his children within a week of marriage as a fraud on his prospective wife's dower rights because he failed to disclose the transfer).

42. Compare *Perlberg v. Perlberg*, 18 Ohio St. 2d 55, 247 N.E.2d 306 (1969) (in the absence of a right to dower existing prior to marriage and a showing of actual fraud, nondisclosure of a transfer of real property one day before marriage will not constitute

disclosure could be viewed as a fraudulent inducement to marry.⁴³ The appropriate remedy was an annulment.⁴⁴ The doctrine, however, did not develop in this way. Instead, the remedy focused on undoing the transfer instead of the marriage. The doctrine developed independently of common-law dower and the elective share system and frequently failed to be integrated into elective share law. Consequently, the relief granted was sometimes broader than that given by the forced share.⁴⁵

As a rule, elective or forced share statutes guarantee the surviving spouse a portion of the property owned by the decedent at death regardless of when it was acquired.⁴⁶ The majority of forced share statutes do not reach property that the decedent transferred without consideration to a third person either before or during marriage,⁴⁷ even if the decedent retained

constructive fraud) with *Chandler v. Hollingsworth*, 3 Del. Ch. 99, 111, 119 (1867) (stating that “[t]he true ground of relief is not *the disappointment of an expectation*, but *fraud upon a legal right*, that is, the right to a marriage without any secret alteration of the circumstances of the parties as they stood at the time of the engagement,” and indicating that only courts of law require legal seisin for the protection of dower).

43. Initially, courts found that constructive fraud based on a failure to disclose was sufficient to invalidate the transfer in whole or in part. Later courts, however, required an actual misrepresentation or a material omission with an intent to deceive and reliance by the other spouse. *See supra* note 41.

44. For a discussion of misrepresentation as a basis for annulment, see Bregy & Wilkinson, *supra* note 33, at 75 (concluding, without explanation, that an “annulment of the marriage would . . . be inappropriate even if annulments or divorces could be granted on this ground, which, of course, they cannot”).

45. Dower and the forced share would only give the surviving spouse a fractional interest in the property. *See Vaughan, supra* note 33, at 182 (with respect to “voidable transfers of property, it may be that the rights of a person not yet married but engaged to be married are superior to the rights possessed as against property transactions occurring after marriage”).

46. Georgia is the only noncommunity property state that does not provide for common-law dower or a forced share. The sole protection afforded the surviving spouse is a year’s support. GA. CODE ANN. § 53-5-1 (1982 & Supp. 1988). *See Note, Preventing Spousal Disinheritance in Georgia*, 19 GA. L. REV. 427 (1985) (urging extension of the equitable property division concept to Georgia probate law in order to recognize each spouse’s presumed contribution to the family wealth); Note, *The Protection of the Surviving Spouse Against Disinheritance: A Search for Georgia Reform*, 9 GA. L. REV. 946 (1975) (providing history and policy of protection against disinheritance). As a rule, community property systems provide the surviving spouse with one-half of the decedent’s property acquired during marriage irrespective of the name in which it is titled. Therefore, antenuptial transfers of nonmarital property are generally not a concern in community property jurisdictions. Consequently, this Article will deal primarily with the rights of surviving spouses in noncommunity property jurisdictions.

47. This was not completely true under common-law dower which generally gave the widow a life interest in one-third of any property of which her husband was seized of an inheritable estate at any time during their marriage. Thus, during marriage, the husband had no power to transfer real property to a third party free of his wife’s dower rights. *See W. MACDONALD, supra* note 33, at 59-64.

certain interests in the property.⁴⁸ Consequently, these statutes enable the decedent effectively to disinherit the surviving spouse and to enjoy economic control over the property. To avoid this circumvention of elective share statutes, courts in some jurisdictions adopted common-law rules that invalidated certain transfers as fraudulent or illusory.⁴⁹ This case-by-case approach frequently produced confusion and inequitable results. In response, some states enacted legislation that subjected certain types of transfers to the right of election based on objective criteria.⁵⁰ These transfers generally involve either a retention of control over the disposition of the transferred property or a beneficial interest in it. Principal among these arrangements are joint tenancies,⁵¹ Totten trusts,⁵² revocable trusts,⁵³ and retained life interests.⁵⁴ This approach was also adopted by the Uniform

48. See, e.g., S.C. CODE ANN. §§ 62-2-201, -202, -205 (Law. Co-op. Supp. 1990) (giving the surviving spouse a right to elect one-third of the property passing under the decedent's will and by intestacy).

49. The history of judicial attempts to prevent disinheritance of the surviving spouse has been discussed extensively. See, e.g., W. MACDONALD, *supra* note 33, at 67-144; Clark, *The Recapture of Testamentary Substitutes to Preserve the Spouse's Elective Share: An Appraisal of Recent Statutory Reforms*, 2 CONN. L. REV. 513, 518-22 (1970); Kurtz, *The Augmented Estate Concept Under the Uniform Probate Code: In Search of an Equitable Elective Share*, 62 IOWA L. REV. 981, 993-1006 (1977); Powers, *Illusory Transfers and Section 18*, 32 ST. JOHN'S L. REV. 193 (1958).

50. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1), (c) (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203 (a) (Purdon Supp. 1991). These statutes served as models for the formulation of the pre-1990 UPC. See U.P.C. § 2-202 comment (1983). The original version of the UPC was approved by the National Conference of Commissioners on Uniform State Laws and the American Bar Association in 1969. Nine states have enacted all or part of pre-1990 UPC § 2-202. See ALASKA STAT. § 13.11.075 (1985); COLO. REV. STAT. § 15-11-202 (1987); IDAHO CODE § 15-2-202 (1979); ME. REV. STAT. ANN. tit. 18-A, § 2-202 (1964 & Supp. 1990); MONT. CODE ANN. § 72-2-705 (1989); NEB. REV. STAT. § 30-2314 (1989); N.D. CENT. CODE § 30.1-05-02 (Supp. 1989); S.D. CODIFIED LAWS ANN. § 30-5A-2 (1984 & Supp. 1991); UTAH CODE ANN. § 75-2-202 (1978).

51. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1)(D) (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203(a)(4) (Purdon Supp. 1991); U.P.C. § 2-202(b)(2)(ii), (iv)(C) (Supp. 1991); U.P.C. § 2-202(1)(iii) (1983).

52. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1)(B) (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203(a)(3) (Purdon Supp. 1991); U.P.C. § 2-202(b)(2)(iv)(B) (Supp. 1991); U.P.C. § 2-202(1)(ii) (1983).

53. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1)(E) (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203(a)(3) (Purdon Supp. 1991); U.P.C. § 2-202(b)(2)(iv)(B) (Supp. 1991); U.P.C. § 2-202(1)(ii) (1983).

54. See, e.g., 20 PA. CONS. STAT. ANN. § 2203(a)(2) (Purdon Supp. 1991); U.P.C. § 2-202(b)(2)(iv)(A) (Supp. 1991). The New York statute does not include transfers with retained life interests in the net estate although these dispositions were discussed by the Bennett Commission which drafted EPTL § 5-1.1. See GOVERNOR AND LEGISLATURE OF NEW YORK, THIRD REPORT OF THE TEMPORARY STATE COMMISSION ON THE MODERNIZATION, REVISION AND SIMPLIFICATION OF THE LAW OF ESTATES: RIGHT OF THE SURVIVING SPOUSE

Probate Code.⁵⁵

Testamentary substitutes, with some exceptions, must be created while the decedent is married to the surviving spouse to be subject to the surviving spouse's right of election under these statutory schemes.⁵⁶ A surviving spouse generally has no right to elect against assets that the decedent transferred to a third person prior to marriage.⁵⁷ Thus, under these statutes, a decedent, whether or not in contemplation of a future marriage, may disinherit a surviving spouse by transferring property to third parties prior to marriage while retaining a beneficial interest in or control over the property until death.

The only restriction on these retained interests or outright transfers prior to marriage was the transfer in fraud of marital rights doctrine. The doctrine frequently prevented the disinheritance of the surviving spouse in situations in which dower and the forced share were inapplicable. However, the difficulties of proving when a transfer was made in contemplation of marriage,⁵⁸ the elements constituting fraud, including

WITH REFERENCE TO NON-TESTAMENTARY TRANSFERS (1964) [hereinafter THIRD REPORT] (Report No. 1.5C). The Third Report expressed its concern "with the situation where . . . the settlor reserves to himself either a *life estate* or a power of revocation or power of control or any combination of these." *Id.* at 124 (emphasis added). In this report, the New York State Bar Association Committee on Trusts and Estates contended that a transfer pursuant to which "practical control of the beneficial enjoyment of the property has been retained during the transferor's lifetime" should be "considered subject to the elective rights of the surviving spouse." *Id.* at 138. Ultimately, the Bennett Commission summarily rejected that view, asserting that "the surviving spouse should *not* be given a right of election merely because 'practical control of the beneficial enjoyment of the property has been retained during the transferor's lifetime.'" *Id.* at 138-39 (emphasis added). This author believes that these transfers should be included in the net estate for elective share purposes. *See also* Kwestel & Seplowitz, *supra* note 1, at 57-59 (arguing that property transferred subject to a retained life interest should be subject to the right of election). The Advisory Committee Report recommended including transfers subject to retained life interests in the net estate if the transfer was effected during marriage. ADVISORY COMMITTEE REPORT, *supra* note 1, at 15-16.

55. U.P.C. § 2-202(b)(2)(ii), (iv)(A)-(C) (Supp. 1991).

56. *See, e.g.*, N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1) (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203(a)(2), (4)-(6) (Purdon Supp. 1991); U.P.C. § 2-202(b)(2)(iv) (Supp. 1991). *But see* Mo. ANN. STAT. § 474.150(1) (Vernon Supp. 1990); Wis. STAT. ANN. § 861.17(1)(a) (West 1991); U.P.C. § 2-202(b)(2)(i)-(iii) (Supp. 1991).

57. *See, e.g.*, *In re Estate of Scheiner*, 141 Misc. 2d 1037, 1038, 535 N.Y.S.2d 920, 921 (1988) (stating that United States Treasury bonds were exempt from the right of election because they were purchased before marriage and also fell within the exclusion of EPTL § 5-1.1(b)(2)).

58. For a discussion of whether a decedent was required to consider the possibility of marriage at some future time or to be engaged to a specific person, see Higgins v. Higgins, 219 Ill. 146, 76 N.E. 86 (1905). Most jurisdictions required an actual engagement. Jarvis v. Jarvis, 286 Ill. 478, 122 N.E. 121 (1919); Beechley v. Beechley, 134 Iowa 75, 108 N.W. 762 (1906).

whether reliance on the prospective spouse's statements and misrepresentation or merely a failure to disclose were necessary,⁵⁹ and any exceptions to the rule,⁶⁰ produced inconsistent results and much litigation. For the same reasons that the fraud on the surviving spouse test used to evaluate transfers made during marriage proved ineffective and unsatisfactory in the elective share area,⁶¹ the doctrine as applied to premarital transfers also has serious defects. The doctrine certainly is not based on the objective criteria that are the product of the more progressive elective share statutes such as the UPC, as well as the Pennsylvania Estates Act and the New York Estates, Powers and Trusts Law (EPTL).⁶² Despite its shortcomings, the doctrine in its modern incarnations has been codified in several states⁶³ and continues to be applied under the common law in numerous jurisdictions. Theoretically, whether the decedent retained any control over or interest in the property at death

59. Courts disagreed as to whether reliance and actual misrepresentation, rather than nondisclosure, were necessary elements of this cause of action. See *infra* note 123.

60. Some courts upheld transfers if the beneficiaries were children of a prior marriage. See, e.g., *Trabbic v. Trabbic*, 142 Mich. 387, 105 N.W. 876 (1905) (commending prospective husband for providing for children where antenuptial transfers consisted of part of a large estate); *Perlberg v. Perlberg*, 18 Ohio St. 55, 247 N.E.2d 306 (1969) (recognizing that transferring real property to children of a prior marriage may be morally and socially meritorious); *King v. Cotton*, 2 P. Wms. 674, 24 Eng. Rep. 912 (1732); *Hunt v. Matthews*, 1 Vern. 408, 23 Eng. Rep. 549 (1686).

61. See THIRD REPORT, *supra* note 54, at 123 (finding "that the case law has been confusing and that the expectancy of the surviving spouse has not been adequately protected"); W. MACDONALD, *supra* note 33, at 3-19, 67-144 (discussing judicial confusion, increased litigation, and tests applied by the courts to *inter vivos* transfers).

62. N.Y. EST. POWERS & TRUSTS LAW § 5-1.1 (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203 (Purdon Supp. 1991); U.P.C. § 2-202 (Supp. 1991); U.P.C. § 2-202 (1983).

63. See, e.g., MO. ANN. STAT. § 474.150(1) (Vernon Supp. (1990) (subjecting "[a]ny gift made by a person, whether dying testate or intestate, in fraud of the marital rights of his surviving spouse" to the right of election); WIS. STAT. ANN. § 861.17(1)(a) (West 1991). Wisconsin codified the transfer in fraud of marital rights doctrine in 1969 to apply to dispositions of property both during marriage and "in anticipation of marriage for the primary purpose of defeating the rights of the surviving spouse." Originally, the statute insulated transfers for the benefit of issue by a prior marriage that were made before marriage or within one year after marriage from attack as a fraudulent property arrangement. WIS. STAT. ANN. § 861.17(2) (repealed 1985). The statute did not define "anticipation of marriage;" therefore, it was unclear whether pre-engagement transfers were subject to challenge. The purpose of the statute was "to prevent depletion of the probate estate at the expense of the surviving spouse." *Id.* § 861.17 comment. In noting that the committee proposing the legislation considered and rejected the Pennsylvania and New York approaches, the comment recognized that "the test of 'primary purpose' [may be] . . . difficult of proof, [but] it has the advantage of being familiar." *Id.* See *infra* notes 129-30.

under a fraud doctrine is largely immaterial.⁶⁴ In this way, the doctrine is also inconsistent with the approach of the UPC, the EPTL, and the Pennsylvania statute which include, for the most part, only testamentary substitutes in the augmented or net estate.⁶⁵

New York, Pennsylvania, and UPC states, although enacting objective criteria to determine the includibility of testamentary substitutes created during marriage in the net estate, have not dealt explicitly with whether transfers (outright or otherwise) made prior to marriage and in fraud of marital rights are valid.⁶⁶ Clarification is therefore needed to determine the validity of such premarriage fraudulent transfers under these statutory schemes and to determine whether the doctrine, if still viable, applies to outright as well as retained interest transfers. More importantly, analysis of whether the standards formulated by these statutes should apply to all antenuptial transfers, such as the situation presented by Case 1, regardless of whether they were made in contemplation of marriage, is critical. As the UPC and state elective share statutes extend the protection given to the surviving spouse by including more post-wedding transfers within the net estate,⁶⁷ decedents will have an incentive to make such conveyances prior to marriage.⁶⁸ State legislatures, model codes, and commentators have not dealt with this problem in a comprehensive or objective way.⁶⁹ Many antenuptial transfers

64. Nevertheless, a review of the cases applying the doctrine reveals that the vast majority involve some form of retained interest in or power over the transferred property.

65. As used in this Article, the terms "augmented estate" and "net estate" refer to property that is subject to the surviving spouse's right of election.

66. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 5-1.1 (McKinney 1981); 20 PA. CONS. STAT. ANN. § 2203 (Purdon Supp. 1991); U.P.C. § 2-202 (1983). This is also true with respect to the testamentary substitutes enumerated in § 2-202(b)(2)(iv) of the revised UPC.

67. The trend in the law has been to expand the protection afforded to the surviving spouse. See *supra* note 50 and accompanying text.

68. Compare the discussion relating to the motives in acquiring insurance in U.P.C. § 2-202 comment (Supp. 1991) with U.P.C. § 2-202 comment (1983).

69. The articles dealing with the doctrine of antenuptial transfers in fraud of marital rights have examined its application critically, but have not suggested replacing it with a more objective approach as is recommended here. See, e.g., W. MACDONALD, *supra* note 33, at 354-65 (recognizing antenuptial transfers as a separate category); Bregy & Wilkinson, *supra* note 33, at 62-63, 74-76 (approving the decision in *Kirk v. Kirk*, 340 Pa. 203, 16 A.2d 47 (1940), which made the requirements for proving fraud more stringent); Lowe, *supra* note 33 (discussing the doctrine in Missouri); Vaughan, *supra* note 33 (discussing the doctrine in New York). Macdonald urged applying a "reasonableness" test to determine whether a prenuptial transfer is subject to the right of election. See W. MACDONALD, *supra* note 33, at 364-65. Recently, however, the Advisory Committee recommended the inclusion of revocable testamentary substitutes in the net estate regardless of whether they were created before or during marriage. ADVISORY COMMITTEE REPORT, *supra* note 1, at 15-16.

continue to be evaluated independently of marital transfers on the assumption that different standards are appropriate to determine whether such dispositions should be subject to dower or the right of election.⁷⁰

Separate treatment of testamentary substitutes created before and during marriage does not promote the policies of elective share statutes. To the extent the surviving spouse requires property for support, whether the testamentary substitute was created prior to marriage should be immaterial. The other justification for the elective share, the partnership theory of marriage, should not produce a different conclusion. Antenuptial testamentary substitutes should be considered partnership assets as long as the decedent's economic control over or benefit from the property continued until or close to death.⁷¹ These assets were just as available to the decedent in terms of disposition and benefit during marriage as they would have been if the decedent created them during the marriage. In effect, by reserving an interest in or power over the transferred property, the decedent brought the property into the partnership. This view accords with the total partnership theory of marriage.

The fact that property conveyed prior to marriage may have been derived from a prior spouse or have been intended for the children of a prior marriage does not justify immunizing it from any claim by the surviving spouse. It merely means that in Case 1, Horace's children may take less from the trust and the house subject to a retained life estate in recognition of Wanda's contribution to the second marriage. (If Horace

70. See, e.g., *Hoeffner v. Hoeffner*, 389 Ill. 253, 59 N.E.2d 684 (1945) (indicating that premarriage transfers are scrutinized more strictly); *Toman v. Svoboda*, 4 Ill. App. 3d 148, 280 N.E.2d 499 (1972) (articulating a different rule for premarriage and marital transfers); *Holmes v. Holmes*, 3 Paige Ch. 363, 364 (N.Y. Ch. 1832) (distinguishing "cases of underhand dealing and secret conveyances made in contemplation of marriage [as] depend[ing] upon an entirely different principle"); W. MACDONALD, *supra* note 33, at 181 (stating that "separate treatment seems warranted" for antenuptial transfers). Support also exists for limiting the surviving spouse's right of election not only to testamentary substitutes created during marriage, but also to property owned by the decedent that was acquired during marriage. See, e.g., OKLA. STAT. ANN. tit. 84, § 44(B) (West 1990) (providing for a right of election in one-half of the "property acquired by joint industry of the husband and wife during coverture"); UTAH CODE ANN. §§ 75-2-201, -202 (1978 & Supp. 1990); Oldham, *supra* note 1, at 246. This view seems to reflect the influence of those who favor applying an equitable distribution approach to the elective share area.

71. See, e.g., I.R.C. § 2036(a)(1) (1986) (including property in the gross estate for estate tax purposes where the decedent transferred property for less than adequate and full consideration in money or money's worth and retained "for his life or for any period not ascertainable without reference to his death or for any period which does not in fact end before his death — (1) the possession or enjoyment of, or the right to income from, the property"); U.P.C. § 2-202(b)(2)(iv) (Supp. 1991) (including property transferred subject to a retained interest or power held within two years of death).

and Fran wanted their children to receive the full amount, then Fran should have left her property to Horace in trust, Horace should have made outright gifts to the children, and Wanda should have waived her right to elect against the antenuptial transfers.) The traditional approach ignores Wanda's contribution to her marriage with Horace because the bulk of Horace's assets were held as premarriage testamentary substitutes. There is no evidence to indicate that Wanda's performance in Case 1 was any less valuable than Hubert's in Case 2 or Horatio's in Case 3. Furthermore, assuming that Wanda has not consented to the transfers, the legitimate expectations of these three surviving spouses do not differ. The proper approach is to balance the interests of the surviving spouse and the children of a prior marriage in testamentary substitutes created prior to marriage, rather than using an exclusionary approach against the spouse.⁷²

II. THE ORIGINS OF THE DOCTRINE IN ENGLISH LAW

The doctrine that invalidated antenuptial transfers in fraud of marital rights originally arose in England where it was applied solely to transfers by prospective wives.⁷³ The justification for the doctrine was twofold: to compensate the husband for his obligation to assume his wife's debts and to support his wife during marriage. Thus, the legal rights protected by this doctrine were primarily rights that existed during the wife's life rather than in her estate.⁷⁴ As the doctrine developed in England, it was not applied to prenuptial transfers by the prospective husband.⁷⁵

The apparent reason for the nonapplicability of the doctrine to antenuptial conveyances by a prospective husband was that a wife had no legal rights to her husband's property during his life and as a practical matter, none at his death. First, a woman had no right to enjoy or control her husband's property during marriage although she certainly had a right to be supported by him. The right of support was never recognized by the English courts as providing a sufficient basis for invalidating an antenuptial transfer by the husband. Second, a wife had

72. This balancing test was also recognized in cases applying the doctrine of antenuptial conveyances prior to marriage. For example, in *Ward v. Ward*, 63 Ohio St. 125, 57 N.E. 1095 (1900), the court recognized a father's obligation to his children, but held that he was not allowed to transfer property during his engagement without notice to his prospective wife.

73. See *infra* note 80.

74. Consequently, a husband was generally required to bring an action to set aside the conveyance during marriage, rather than after his wife's death. See, e.g., *Loader v. Clarke*, 2 Mac. & G. 384, 42 Eng. Rep. 148 (1850).

75. See *Chandler v. Hollingsworth*, 3 Del. Ch. 99, 106 (1867) (discussing the development of the doctrine in England).

an inchoate dower interest in her husband's real property which he owned as an inheritable estate during marriage. As a general rule, however, a woman waived her dower interest before marriage by jointure in which the prospective husband agreed to a property settlement for the benefit of his intended wife. Due to the widespread practice of jointure, premarriage transfers by the husband did not deprive the woman of any legal rights because she essentially did not acquire any by marriage.⁷⁶ This was not the case for the husband.

As formulated by the English courts, the doctrine of antenuptial transfers in fraud of marital rights invalidated conveyances made by a woman with the intent to keep her assets from her prospective husband after a treaty of marriage or an engagement and before the actual marriage. In some cases, the motive for the transfer was to prevent the prospective husband from obtaining control of the property. In other situations, the primary purpose was to provide for the children of a prior marriage. Thus, the timing of the transfer, the woman's intent, and the husband's knowledge of the conveyance⁷⁷ were critical in determining whether the transfer was fraudulent.

Misrepresentation or failure to disclose a property disposition is an understandable element in an action for actual or constructive fraud. Yet, why was the timing such a critical factor both in England and later in the United States? Theoretically, a woman should have been able to give her property away prior to the wedding without infringing on her husband's rights because a man had no legal interest in a woman's property until marriage. Nevertheless, a man often relied on the extent of his intended wife's assets in agreeing to the marriage,⁷⁸ as well as relying on property settlements that accompanied a treaty of marriage. A man might decide not to go through with the marriage or might agree to a less generous property settlement if he knew of the antenuptial conveyance. In addition, a transfer that occurred prior to an engagement generally would not violate a spouse's dower or other marital rights on the theory that no imminently prospective rights existed that would allow an inference of fraudulent intent.⁷⁹

76. *Id.* at 116 (noting the practice in England of jointure in which a husband would make a settlement upon his intended wife that, if reasonable, was the equivalent of and would bar dower).

77. Knowledge on the part of the husband could be found if he had reason to know that his wife transferred property. See *Wrigley v. Swainson*, 3 De G. & Sm. 458, 64 Eng. Rep. 560 (1859). In contrast, some United States courts required that the husband not only know, but also consent to the transfer. See *infra* note 89.

78. An extreme example of a case in which the husband married his wife solely for her property can be found in *Strathmore v. Bowes*, 1 Ves. jun. 22, 30 Eng. Rep. 211 (1789), *aff'd sub nom. Bowes v. Bowes*, 6 Brown 427, 2 Eng. Rep. 1178 (1797).

79. See *Melenky v. Melen*, 233 N.Y. 19, 134 N.E. 822 (1922) (upholding the transfer

The English courts showed a marked reluctance to apply the doctrine to invalidate a woman's antenuptial transfers despite a strong endorsement of the doctrine in theory.⁸⁰ In the leading case of *Strath-*

of property to a son because it was made before the second marriage was contemplated and for adequate business considerations); *Bliss v. West*, 58 A.D. 71, 11 N.Y.S. 374 (1890) (upholding a transfer of real estate in trust for the benefit of the children of a prior marriage with a reserved life estate made before the engagement and possibly before the grantor met his future wife, although he represented to her that he owned the property). In addition, an engaged couple is considered to be in a confidential relationship that may provide a basis for a duty to disclose. See *Pierce v. Pierce*, 71 N.Y. 154, 157 (1877).

80. A survey of a large number of English cases involving antenuptial transfers indicated that generally, the prospective wife retained an interest and the court upheld the transfer. See *Downes v. Jennings*, 32 Beav. 290, 297, 55 Eng. Rep. 114, 117 (1863) (holding that a wife's transfer of a long-term leasehold seven weeks before marriage for her benefit for life and then for the benefit of other relatives was invalid as a fraud on the husband's marital rights because he believed that "this property of his intended wife would enable him to support her"); *Loader v. Clarke*, 2 Mac. & G. 382, 387, 42 Eng. Rep. 148, 150 (1850) (stating the general rule that if the wife's antenuptial transfer of her annuity to trustees for her benefit for life and then to her children had been concealed from her husband it would have been fraudulent, but because the husband must have known of the trust, waiting until his wife's death to recover stock constituted a waiver); *Grazebrook v. Percival*, 14 The Jurist 1103 (1850) (dicta that if the husband had been alive, he could have recovered possession of stock that his wife, shortly before marriage, had transferred in trust for herself without his knowledge); *Wrigley v. Swainson*, 3 De G. & Sm. 458, 64 Eng. Rep. 560 (1849) (upholding a wife's transfer of money during the engagement to trustees for her benefit for life, remainder to her children or, if none, as she appointed, when her husband knew before the marriage that his wife intended to dispose of the bulk of her property despite his ignorance of the terms of the actual trust and had not consented to any settlement); *Griggs v. Staplee*, 2 De G. & Sm. 572, 64 Eng. Rep. 255 (1848) (upholding a wife's prenuptial transfer of money in trust for her benefit in the absence of evidence that a transfer occurred after the engagement and that the husband did not know of the disposition); *Taylor v. Pugh*, 1 Hare 608, 66 Eng. Rep. 1173 (1842) (upholding a wife's transfer of property in trust for the benefit of herself and her children because her husband persuaded her to live with him before marriage, thereby precluding her from obtaining a settlement from him); *England v. Downs*, 2 Beav. 522, 48 Eng. Rep. 1284 (1840) (stating the general rule and rejecting the exception when property is derived from a prior husband and given to the children, but upholding the wife's transfer two months prior to marriage of all of her property in trust for herself and her children because of insufficient evidence to show that the transfer occurred after a treaty of marriage or that the husband had no notice); *St. George v. Wake*, 1 My. & K. 610, 39 Eng. Rep. 812 (1833) (upholding a wife's transfer of a small portion of her property to her sister one month before marriage because the evidence indicated that her husband had knowledge of the transfer prior to the marriage, he had no money of his own, and he did not make a settlement with his prospective wife); *Goddard v. Snow*, 1 Russ. 485, 494-95, 38 Eng. Rep. 187, 191 (1826) (invalidating a wife's transfer of property 10 months before marriage to a trust for her benefit on the basis of the rule that "if a woman, after contemplating marriage with an individual, make a settlement which she conceals from her intended husband, that settlement is not good as against him"); *De Manneville v. Crompton*, 1 V. & B. 354, 35 Eng. Rep. 138 (1813) (upholding a wife's

more v. Bowes,⁸¹ the court upheld a transfer made days before the wedding on the ground that the woman was engaged to another man at the time of the transfer and had informed him of the conveyance. Despite the woman's failure to disclose the disposition of her property to the man she actually married and the fact that he married her solely for her property, the court found that the disposition was not a fraud on his marital rights because he was not engaged to her at the time of the transfer. Thus, the statement of the rule was essentially dicta. Nevertheless, *Strathmore* has been cited repeatedly in support of this doctrine both in England and in the United States, particularly its justification as articulated by Lord Thurlow:

The law conveys the marital rights to the husband, because it charges him with all the burdens, which are the consideration, he pays for them; therefore it is a right, upon which fraud may be committed. Out of this right arises a rule of law, that the

transfer of property to a trust for her benefit in which her husband took a partial contingent interest without her prospective husband's knowledge because the wife never represented to him that the note would not be canceled); *Ball v. Montgomery*, 2 Ves. jun. 191, 194, 30 Eng. Rep. 588, 590 (1793) (dicta that "[i]f a woman previously to marriage conveys her property without the privity of the intended husband, it will be fraud"); *Strathmore v. Bowes*, 1 Ves. jun. 22, 30 Eng. Rep. 211 (1789), (upholding transfer by a wife 10 days prior to marriage while engaged to another man to whom she disclosed a settlement for the benefit of herself and her children), *aff'd sub nom. Bowes v. Bowes*, 6 Brown 427, 2 Eng. Rep. 1178 (1797); *Blanchet v. Foster*, 2 Ves. sen. 265, 28 Eng. Rep. 171 (1751) (upholding a wife's transfer of a bond to her aunt just before the marriage because she gave it for valuable consideration); *King v. Cotton*, 2 P. Wms. 674, 24 Eng. Rep. 912 (1732) (upholding transfer by a wife for the benefit of herself and her children by a former marriage because her husband did not make a jointure for the benefit of his prospective wife, the transfer occurred before the engagement, and it was reasonable for her to want to provide for her children); *Carleton v. Earl of Dorset*, 2 Vern. 17, 23 Eng. Rep. 622 (1686) (invalidating transfer by a wife prior to marriage because the wife was allegedly in debt and told creditors to sue her prospective husband); *Hunt v. Matthews*, 1 Vern. 408, 23 Eng. Rep. 549 (1686) (upholding an antenuptial transfer of most of a wife's assets to a trustee for the benefit of her daughter because the husband suppressed it after marriage and the desire to provide for children of a prior marriage was justifiable); *Thomas v. Williams*, 2 Ch. Rep. 79 (1672-73) (no fraud on marital rights when during her engagement, a wife assigned to her brother money her mother had promised her from her father's estate which belonged to her brother); *Lance v. Norman*, 2 Ch. Rep. 79 (1672-73) (invalidating the transfer of money by a wife to her brother one day before marriage without her husband's knowledge, but ordering the husband to pay the wife an annuity for her separate maintenance); *Howard v. Hooker*, 2 Ch. Rep. 81 (1672-73) (setting aside a wife's transfer of her interest in her first husband's estate to her daughter because her husband relied on her ownership of the property in agreeing to jointure).

81. 1 Ves. jun. 22, 30 Eng. Rep. 211 (1789), *aff'd sub nom. Bowes v. Bowes*, 6 Brown 427, 2 Eng. Rep. 1178 (1797).

husband shall not be cheated, on account of his consideration. . . . A conveyance by a wife, whatsoever may be the circumstances, and even upon the moment before the marriage, is *prima facie* good; and becomes bad only upon the imputation of fraud. If a woman during the course of a treaty of marriage with her makes without notice to the intended husband a conveyance of any part of her property, I should set it aside, though good *prima facie*, because affected with that fraud.⁸²

As this excerpt from *Strathmore* indicates, a woman's fraudulent intent could be gleaned from the circumstances: if the transfer occurred during the engagement and the prospective husband did not have notice, the transfer theoretically would be fraudulent on a constructive fraud theory.⁸³

III. THE DOCTRINE IN THE UNITED STATES

A. *The Early Period*

A broad interpretation of the doctrine took root early in the United States. In *Manes v. Durant*,⁸⁴ the court stated that “[i]n general, it is not questioned that a voluntary conveyance made by a woman in contemplation of marriage, without the knowledge of the intended husband, will be set aside as a fraud on the marital rights.”⁸⁵ By invalidating a woman's antenuptial transfer of her slaves and personality in trust for the benefit of various relatives including her surviving children, the court rejected any exception to the doctrine when the disposition favored children.⁸⁶ The expansive scope of this doctrine was reached by likening

82. *Id.* at 28, 30 Eng. Rep. at 214. The court, in *Goddard v. Snow*, 1 Russ. 495, 38 Eng. Rep. 191 (1826), interpreted Lord Thurlow's opinion in *Strathmore v. Bowes* to apply to settlements of property by a woman in which she “reserv[es] to herself the dominion over it.”

83. Nevertheless, despite this broad statement of the general rule, most of the English courts found the antenuptial transfers in question to be valid. Parenthetically, the court in *Ward v. Ward*, 63 Ohio St. 125, 127, 57 N.E. 1095, 1096 (1900), misstated the actual practice in England when it observed that antenuptial “conveyances by the wife are uniformly held invalid.”

84. 2 Rich. Eq. 404 (S.C. 1842).

85. *Id.* at 405-06. Thus, an undisclosed transfer made for valuable consideration will not be deemed to be fraudulent. *See, e.g.*, *Gregory v. Winston's Adm'r*, 64 Va. 102, 123 (1873) (citing *Blanchet v. Foster*, 2 Ves. sen. 265, 28 Eng. Rep. 171 (1751)); *Fletcher v. Ashley*, 47 Va. 332, 338 (1849).

86. *Manes*, 2 Rich. Eq. at 406. Some courts recognized, at least in theory, that a premarriage transfer for the benefit of children of a prior marriage could vitiate the fraud. *See supra* note 60.

a prospective husband to a bona fide purchaser for value. As in England, the justifications for the doctrine in the early cases in the United States were the husband's obligation to support his wife and to assume her pre-existing debts.⁸⁷ The liability for these debts, it was believed, gave him an interest in his wife's property.⁸⁸ Transfers of property by a prospective wife without the intended husband's knowledge or consent⁸⁹ deprived him of more than an expectancy in her property.⁹⁰ This purported justification for the doctrine is subject to question. If the intended

87. See, e.g., *Arnegaard v. Arnegaard*, 7 N.D. 475, 480, 75 N.W. 797, 798 (1898) ("An agreement for marriage at common law was, in effect, an agreement for a sale by the prospective wife to the prospective husband of all her personal property, and the transfer to him of her right to possession of all her real estate, on condition of the assumption by him of all her debts."); *Gregory v. Winston's Adm'r*, 64 Va. 102, 123-24 (1873). The husband's acquisition of his wife's property upon marriage, however, was also justified by his obligation to maintain her. See *Chandler v. Hollingsworth*, 3 Del. Ch. 99, 106 (1867). In such a case, it could be argued that the husband should be able to reach the transferred property to the extent needed to maintain his wife.

88. See *Ramsay v. Joyce*, 1 McMul. Eq. 236, 252 (S.C. 1841) ("it is the familiar law, that the husband is regarded as a purchaser of the wife's property, and that marriage is a valuable consideration"). See also *Manes v. Durant*, 2 Rich. Eq. 404 (S.C. 1842).

89. Courts disagreed as to whether constructive knowledge, actual knowledge, or consent was necessary to prevent an antenuptial transfer from being fraudulent. See *Spencer v. Spencer*, 56 N.C. 404, 409 (1857) ("[t]here must be a knowledge of and assent to the particular deed to give it the effect of barring the rights of the husband"). Compare *Fletcher v. Ashley*, 47 Va. 332, 339 (1849) (Brooks, J., concurring) (knowledge is a sufficient bar) with *Johnson v. Peterson*, 59 N.C. 12, 14 (1860) (conveyance by woman on the eve of her marriage constitutes fraud upon intended husband's rights unless he had full knowledge of the transaction and freely assented to it). Of course, establishing knowledge or consent many years after a transfer or after the grantor's death was extremely difficult and was responsible for many lawsuits yielding frequently inconsistent results. See also *Poston v. Gillespie*, 58 N.C. 258 (1859), in which the court invalidated a woman's transfer of land, a slave, and personality for the benefit of herself and her son from a prior marriage on the ground that he did not consent even though she notified her intended husband of the transfer. The court indicated, in dicta, that a transfer is not fraudulent if it is made by a woman before the engagement and if the man has notice at the time of the transfer because a contract to marry did not exist at that time. The court noted that any property transferred without consent will breach the contract, although equity will not specifically enforce a contract to marry. Prior to courtship or contemplation of marriage, a woman was free to dispose of her property without informing the man she later married. *Id.* at 262. The controversy over whether knowledge or consent is necessary to sustain the transfer continued after the passage of the Married Women's Property Acts. See *Taylor v. Taylor*, 199 N.C. 197, 148 S.E. 171 (1929). For a case finding constructive notice sufficient to prevent fraud, see *McClure v. Miller*, 1 Bail. Eq. 107, 109 (S.C. 1830) ("There was at least enough to put [the husband] upon inquiry."). In *Moore v. Moore*, 15 Ill. 2d 239, 243, 154 N.E.2d 256, 258 (1958), however, the court held that recording a deed prior to marriage does not constitute constructive notice to a prospective wife.

90. *McClure*, 1 Bail. Eq. at 109 (invalidating a wife's transfer prior to marriage for the benefit of a child when her husband, during marriage, satisfied a premarriage debt incurred by his wife for part of the conveyed property).

husband's interest in the property was based on his liability for the woman's debts, then the prospective wife's transfers of property should have been overturned only to the extent of her debts.⁹¹ The husband's interest in the transferred property, however, was generally either total or equal to the share that he could expect if his wife owned the property at her death. His share could also be justified on the ground that the husband was obligated to support his wife. Generally, no correlation was made between a husband's interest in the transferred property and the amount needed for support.

Strictly speaking, none of these reasons adequately supports an action for fraud because the property was conveyed in fraud of future property rights.⁹² Nevertheless, the view of marriage as a sales contract pursuant to which the husband undertakes onerous obligations in exchange for the purchase of his wife's property explains why equity afforded a remedy to what otherwise would appear to be the loss of a mere expectation. The basis for finding fraud in the case of a disappointed expectation was elucidated by Chief Judge Ruffin in *Logan v. Simmons*:⁹³

[T]he husband shall not be cheated, on account of his consideration. Now, the rights . . . are not present rights, that is, existing at the time of the conveyance; for, a fraud on rights of that kind, the common law would redress. They are prospective rights — those that the husband expects to enjoy upon the contemplated marriage by the law of the land. A husband, being bound to pay his wife's debts and to maintain her during coverture, and being chargeable by the law with the support of the issue of the marriage, and bound by the ties of natural affection also to make provision for the issue, it is in the nature of things, as a matter of common discretion, that a woman's apparent property should enter materially, if not essentially, into his inducements for contracting the marriage, and incurring those onerous obligations.⁹⁴

91. See *id.* (dicta that if a woman fraudulently concealed a debt and voluntarily conveyed "her whole property, or so much of it as to leave her unable to satisfy the debt, the husband, if he should be compelled to pay the debt, might be allowed to stand in the place of the creditors and avoid the conveyance, although he knew of it before the marriage;" in that case, "the actual fraudulent intention must be established").

92. An analogy can be drawn to a transfer in fraud of future creditors; however, this should only apply if the prospective spouse is insolvent at the time of the transfer. See *Rivers v. Thayer*, 7 Rich Eq. 136, 156-57 (S.C. 1855).

93. 38 N.C. 487 (1845).

94. *Id.* at 495. This was also the case in *Spencer v. Spencer*, 56 N.C. 404 (1857), in which an engaged woman transferred her slaves a few days before her marriage, reserving a life estate. Although her husband had the right to possess the slaves by virtue

In *Logan*, on the day before her second marriage and without informing her prospective husband, a woman deeded two slaves to her son, reserving the first living child one of them might have. After the marriage, the slaves, who represented the bulk of the woman's property, resided with the couple until the wife's death ten years later. At that time, the son claimed them based on the deed. There was no evidence that the wife incurred any debts prior to her second marriage. Indeed, the evidence indicated that her husband, who was somewhat impecunious, married her for her modest property.⁹⁵ The husband did not have any prospective rights of which he was defrauded. Because his intended wife was solvent, he did not need her property to discharge her debts. She essentially retained a *de facto* life estate in the transferred slaves during their marriage. The husband consequently enjoyed the benefits of her property and presumably could have used the slaves to discharge his support obligation.⁹⁶ Nevertheless, the court invalidated the earlier transfer as a fraud on the husband's marital rights. The court did not include inheritance rights among the rights of which his prospective wife allegedly defrauded him.⁹⁷ Thus, the analytical difficulty in supporting this result is apparent. The court was so offended by the prospective wife's concealment of the transfer from her intended husband who relied on her apparent ownership of the slaves in proceeding with the marriage that it was not troubled by reaching a result that could not be supported by its articulated justifications.⁹⁸

A line of cases that impressed a constructive trust to remedy fraud in preventing a decedent from executing a will and thereby protecting the expectancy interest of the beneficiary provides a useful analogy. In

of the life estate, the court invalidated the transfer as fraudulent on behalf of the deceased husband's creditors. The court did not appear troubled by the fact that the wife's property was used to discharge the husband's debts and was also employed for the couple's support during their marriage.

95. *Logan*, 38 N.C. at 488. It appeared that the husband, whose older wife was in failing health, owned one mare.

96. See *Waller v. Armistead's Adm'rs*, 29 Va. 11, 14 (1830) (prospective wife transferred slaves and the proceeds of a sale of real estate to her brother and his infant son on the morning of her marriage without the knowledge of her fiancé; however, the husband lost his rights because he died during his wife's lifetime).

97. See *supra* note 94 and accompanying text.

98. Interestingly, the court noted that, unlike the situation in England, jointure was not common in the United States and in fact, the man had not settled any property on his prospective wife. The court concluded that a more stringent rule of concealment should apply in the United States because settlements in England are preceded by investigations and the expectations of the parties are limited to the property secured by the agreement. *Logan v. Simmons*, 38 N.C. 487, 497, 501 (1845). This, of course, does not necessarily follow because a greater expectation may exist that property will not be transferred without notice when a settlement has, in fact, been made.

Pope v. Garrett,⁹⁹ the court imposed a constructive trust on property that passed by intestacy to the decedent's heirs in favor of the beneficiary named in an unexecuted will. The actions of several of the presumptive heirs prevented the decedent from executing the will. Despite the fact that the beneficiary had only an expectancy and no existing property rights of which she could be defrauded at the time of the heirs' acts, the court held that a constructive trust was appropriate to remedy the wrongful act and to prevent unjust enrichment.¹⁰⁰ Thus, the heirs were found to have defrauded the intended beneficiary of an expectancy in the decedent's estate, just as the decedent, in a case applying the doctrine of antenuptial transfers in fraud of marital rights, was found to have violated an expectancy of the prospective spouse. Interestingly, marital rights cases generally do not speak in terms of preventing unjust enrichment. In the premarital transfer cases, the decedent was unjustly enriched by the benefits obtained from the marriage (for which the appropriate remedy should be annulment) or by a benefit in the unrestricted disposition of property.

B. The Period Following the Passage of the Married Women's Property Acts

The major justifications for the common-law doctrine disappeared after the passage of the Married Women's Property Acts in the mid-nineteenth century.¹⁰¹ The doctrine, however, did not die, but was transformed to protect the rights of the surviving spouse in the deceased spouse's estate.¹⁰² Thus, the focus of the doctrine changed markedly in this country, although this change is frequently not acknowledged by the courts. Protection of the surviving spouse's rights meant that the doctrine could apply equally to men and women. Suits challenging the conveyance could be brought after the transferor's death without being subject to the defense of laches.¹⁰³

99. 147 Tex. 18, 211 S.W.2d 559 (1948).

100. *Id.* at 23, 211 S.W.2d at 561. In reaching its decision, the court relied on Professor Scott's treatise on trusts which notes a conflict of authority on this point.

101. One commentator argues that the general rule enunciated in *Strathmore v. Bowes*, which invalidated a settlement of property as a fraud on marital rights, is "obsolete in consequence of the Married Women's Property Acts." Shannon, *The Countess of Strathmore versus Bowes*, 1 CANADIAN B. REV. 425, 427 (1923). *But see* Youngs v. Carter, 10 Hun. 194, 198 (N.Y. 1877) (disagreeing with the principle that applied the doctrine only to transfers by prospective wives).

102. The transformation of the doctrine into a rule preventing spousal disinheritance was facilitated by earlier American courts applying it to invalidate transfers by a prospective husband in fraud of his wife's dower rights. For a discussion of the doctrine's history in the United States, see *supra* note 35.

103. If the husband was defrauded of the right to possess his wife's personality and

The transformation of the doctrine rendered its theoretical underpinnings more problematic. The rationale articulated by the English and early American courts, namely, the right of the husband to manage his wife's realty and dispose of her personality in return for his obligation to support her and pay her debts, was no longer relevant. What supported the application of the doctrine to antenuptial transfers of the wife as well as the husband? What marital rights were defrauded?¹⁰⁴ Initially, some courts spoke of the woman's justifiable expectation of support from her husband that required some assurance that he actually owned property with which to fulfill this obligation. In essence, this expectation is the counterpart of the argument used to justify the application of the doctrine to a woman's antenuptial transfers prior to the passage of the Married Women's Property Acts.¹⁰⁵ Formerly, the husband had the right to consume and manage his wife's property to discharge his obligation to support her. After the passage of the Married Women's Property Acts, each spouse had a reciprocal duty of support. Thus, the wife had the right to expect her husband to use his own funds to support her. As stated by the court in *Brooks v. McMeekin*:¹⁰⁶

[D]oes not the wife have a right to look carefully, to see if he who would wed her has the means essential to support herself and such offspring as a kind Providence may give her? In England, is it not a matter of every-day occurrence that the husband makes settlements upon the intended wife often at the instance of his intended wife's family or friends? . . . If a man cannot be charged with indelicacy in looking into his wife's estate, surely the woman, who at best is so unprotected, may consider, when she is asked to marry a man, whether he has the means to provide for her.¹⁰⁷

to manage her realty during marriage and learned of the transfer after the wedding, but did not challenge it until his wife's death, the defense of laches could apply. See, e.g., *Loader v. Clarke*, 2 Mac. & G. 384, 42 Eng. Rep. 148 (1850). Cf. *Babcock v. Babcock*, 53 How. Pr. 97 (N.Y. 1876) (allowing a wife to sue before her husband's death to protect inchoate dower rights where the prospective husband transferred property to his children three days before the marriage).

104. Even before the passage of the Married Women's Property Acts, use of the doctrine to protect a woman's dower rights was rejected in an action at law in *Baker v. Chase*, 6 Hill 482, 483 (N.Y. 1843), because the "husband was not seised at any time during the coverture." The court also questioned the doctrine's application by a court of equity in *Swaine v. Perine*, 5 Johns. Ch. 482 (N.Y. 1821).

105. The husband's obligation to support his wife and assume her debts gave him a right to her property during marriage. See *supra* text accompanying notes 73-74.

106. 37 S.C. 285, 15 S.E. 1019 (1892).

107. *Id.* at 305, 15 S.E. at 1024.

Like its counterpart, this rationale was often untenable. If the husband retained a life estate, which was frequently the case, the property could be utilized to support his wife; therefore, any antenuptial conveyance generally would not have deprived her of her right to maintenance. Consequently, she could not have been the victim of fraud.

The justification for this doctrine was not limited to the duty of support even in *Brooks v. McMeekin*. The court recognized that, for women as well as men, "marriage is a valuable consideration" pursuant to which she acquires not only the right to comfortable maintenance, but also the right to share in her husband's intestate estate and the right to dower.¹⁰⁸ In invalidating the transfer, the court found that the wife was defrauded of an expectancy much like the expectancy the beneficiary was deprived of in *Pope v. Garrett*.

As the doctrine evolved, the central rights protected were principally those relating to a surviving spouse's interest in the deceased spouse's estate. These rights, for the most part, did not involve either spouse's enjoyment of any property owned by the other during marriage, but rather a dower interest or the expectation of receiving a forced share of the decedent's property at death. With respect to antenuptial transfers of personality, the doctrine presented real problems because each spouse generally had the right during marriage to transfer personal property free of the other spouse's claim. At common law and under the typical forced share statute, neither spouse had a property interest in the other spouse's personal estate during marriage. If a spouse was free to transfer personality during marriage, it would seem that the spouse could make such transfers prior to marriage. Nevertheless, the doctrine was applied to prenuptial transfers of personality in some jurisdictions.¹⁰⁹

The application of the doctrine to antenuptial transfers of realty was less problematic, particularly with respect to transfers by prospective

108. *Id.* at 304, 15 S.E. at 1023. Indeed, the court in *Chandler v. Hollingsworth*, 3 Del. Ch. 99 (1827), found that the wife's right to dower in the United States was extremely critical to her because of the absence of jointure. In fact, the court stated, if anything, the woman, as "the weaker sex," needed greater protection than the man under the doctrine because in return for dower she "surrender[ed] her person, her services, her self-control, her means of self-support; and as to property, far more than the interest she acquires." Although the court indicated that the existence of a marital right in dower obviated any need to value the consideration furnished by the wife, its characterization of the detriment suffered by the woman upon marriage influenced other courts in applying a rule that was protective of the surviving spouse. See, e.g., *Ward v. Ward*, 63 Ohio St. 125, 57 N.E. 1095 (1900). Nevertheless, courts in some jurisdictions began to cut back on the scope of the doctrine by requiring a showing of actual fraud. See, e.g., *Perlberg v. Perlberg*, 18 Ohio St. 55, 247 N.E.2d 306 (1969) (overruling *Ward*); *Kirk v. Kirk*, 340 Pa. 203, 16 A.2d 47 (1940); *Dudley v. Dudley*, 76 Wis. 567, 45 N.W. 602 (1890).

109. See, e.g., *LeStrange v. LeStrange*, 242 A.D. 74, 273 N.Y.S. 21 (1934).

husbands. Although a woman had no interest in a man's property prior to marriage, she did have rights in the form of inchoate dower during marriage. While she could not prevent her husband from transferring his personal property during marriage, any transfer of realty by him without her consent was subject to her dower right. An antenuptial transfer of realty deprived the wife of her inchoate dower interest which attached at the time of coverture. Of course, in jurisdictions that abolished dower and replaced it with a forced share, allowing the overturn of antenuptial transfers of realty raised the same conceptual problems that existed in the case of personality.

Generally, the development of common-law doctrines that invalidated certain transfers during marriage did not make the doctrine any more justifiable. First, the illusory transfer and the reality of the transfer tests only applied to dispositions in which the spouse retained significant powers or did not make a real transfer.¹¹⁰ Particularly in its earlier stages, the antenuptial transfer in fraud of marital rights doctrine invalidated many retained interest transfers that would have been valid under either the illusory transfer or reality of the transfer test.¹¹¹ Curiously, in some jurisdictions, such as Pennsylvania and Ohio, as the protection afforded to the surviving spouse increased under elective share statutes, the scope of the doctrine became more limited with little recognition of the trend in either area or the relationship between the two bodies of law.¹¹² Nevertheless, in many jurisdictions, a premarital transfer often involved a retained life estate which, standing alone, would be valid under the illusory transfer and the reality of the transfer tests.¹¹³ Only the motive for the transfer test, which applied to outright as well as retained interest transfers, was somewhat consistent with the test applied to judge the validity of premarriage dispositions.¹¹⁴ In view of the fact that inheritance rights were protected in both premarriage and marital transfer cases, it is difficult to comprehend these divergent approaches.

110. See *supra* note 49 and accompanying text.

111. See, e.g., *Spencer v. Spencer*, 56 N.C. 404 (1857); *Arnegard v. Arnegard*, 7 N.D. 475, 75 N.W. 797 (1898).

112. For example, in Pennsylvania, the court in *Kirk v. Kirk*, 340 Pa. 203, 16 A.2d 47 (1940), overruled prior decisions that utilized a constructive fraud test to determine the validity of antenuptial transfers. *Kirk* imposed an actual fraud test that required an intent to disinherit the surviving spouse. The decision in *Kirk* was applauded in an article by Bregy & Wilkinson. See Bregy & Wilkinson, *supra* note 33. See also *infra* notes 158-59 (discussing Pennsylvania decisions) and *supra* note 108 (discussing Ohio decisions).

113. See, e.g., *LeStrange*, 242 A.D. at 74, 273 N.Y.S. at 21.

114. Compare *Bodner v. Feit*, 247 A.D. 119, 286 N.Y.S. 814 (1936) with Mo. ANN. STAT. § 474.150 (Vernon Supp. 1990) (transfers made with the purpose of destroying the spouse's right of inheritance deemed to be in fraud of the marital rights of the surviving spouse).

The doctrine of antenuptial transfers in fraud of marital rights protects a prospective spouse from duplicitous, nonconsensual transfers that interfere with inheritance rights, but it has serious limitations as a prototype for an elective share statute. Aside from the difficulties associated with a subjective test and its questionable theoretical foundations, the doctrine generally would not provide Wanda (Case 1) with a remedy because Horace transferred his property prior to their engagement. The only relief it might provide would be with respect to the additions Horace made to the trust after the engagement.

Secondly, the question arises whether the doctrine provides a satisfactory remedy if Horace made the transfers while he was engaged to Wanda. Under a constructive fraud approach in a jurisdiction that recognizes dower, a court would probably invalidate the transfer of the house at least to the extent of Wanda's dower rights. As previously discussed, inchoate dower is considered a right that vests upon marriage.¹¹⁵ If Wanda's interest in the corpus of the trust, absent a transfer, was an expectancy during marriage, she could only invalidate the trust in a jurisdiction that recognized marital rights in the expectation of inheritance.¹¹⁶ With respect to either the transfer of the house alone or both the house and the personality in trust, a court would presume fraud absent any misrepresentation by Horace. Under a constructive fraud theory, the existence of these marital rights would impose a duty on Horace to disclose the dispositions to Wanda during their engagement. This duty particularly existed when Horace transferred almost all of his property even though his children were the beneficiaries.¹¹⁷ A duty to disclose makes sense from a moral viewpoint because it fosters honesty between prospective spouses. Yet, is this duty legally tenable?

First, consider the situation in which Wanda knew that Horace resided in the house and appeared to have some property interest in it. Wanda also knew that Horace received income on a regular basis. Is it reasonable to place on her the burden of ascertaining whether Horace owned the house in fee or whether he received the income from a trust rather than from an account titled in his name? The majority of courts applying a constructive fraud test would probably not impose such a burden on Wanda in the absence of any disclosure by Horace. Instead, they would likely find that she justifiably relied on his apparent property ownership.¹¹⁸ This would seem to be the correct result because it avoids the arguably undesirable result of requiring a prospective spouse to

115. See *supra* note 108 and accompanying text.

116. See *supra* note 109 and accompanying text.

117. See *supra* note 60 and accompanying text.

118. Courts frequently did not require a showing of actual reliance in the constructive fraud area. See, e.g., *Chandler v. Hollingsworth*, 3 Del. Ch. 99 (1867).

scrutinize the partner's property holdings in the absence of a prenuptial agreement. This approach would encourage Horace, who made the transfers, to avoid the application of the doctrine by giving notice to Wanda, who would then be bound even absent a written waiver.¹¹⁹

A more difficult case is presented if Wanda lacked knowledge of the existence of the house and the trust and Horace had no plans of using them as marital property, but intended to keep his retained interest in them solely for his personal benefit. Although Wanda clearly could not have relied on this property in marrying Horace, her ignorance is not necessarily fatal to a claim for constructive fraud.

Interestingly, the court in *Chandler v. Hollingsworth*¹²⁰ was not perturbed by a prospective spouse's lack of knowledge of the assets transferred because she was still entitled to a property right in the real estate upon marriage.¹²¹ This reasoning is problematic. A person can have a property right without knowledge of it. Defrauding a prospective spouse of a future property right that may never become possessory, and is really nothing more than an expectation, however, reveals a logical weakness requiring some reliance. A better approach is to dispense with the fiction of the existence of protectable rights during engagement and the problem of reliance and recognize that Horace's retention of an interest in or power over the property means that he has not cut the strings connecting him to the property. Under this theory, which is based on an estate tax analogy,¹²² Horace's retention of a lifetime ownership interest in the property gives Wanda a right to elect against these existing interests, thereby avoiding the machinations associated with the doctrine.

Wanda would have a much more difficult time recovering in a jurisdiction that requires actual fraud. First, she would have to show that she owned a property interest during the engagement. This would be particularly difficult in a jurisdiction like Ohio after *Perlberg v. Perlberg*.¹²³ Second, after overcoming this hurdle, she would have to prove that Horace transferred the property with an intent to disinherit her. In this case, she would rely on Horace's transfer of virtually all of his assets and his retention of a life estate and power over the principal. Nevertheless, she would have to deal with the issue of whether Horace's primary motive was disinheritance or a desire to benefit his

119. This assumes that the action would be brought in a jurisdiction that does not require consent. See *supra* note 89.

120. 3 Del. Ch. 99 (1867).

121. *Id.* at 110-11.

122. See *supra* notes 21-22 and accompanying text.

123. 18 Ohio St. 2d 55, 247 N.E.2d 306 (1969) (in the absence of a right to dower existing prior to marriage and a showing of actual fraud, nondisclosure of a transfer of real property one day before marriage will not constitute constructive fraud).

children even though the latter incorporates an intent to diminish Wanda's share. Lastly, Wanda would have to demonstrate that Horace misrepresented the title to the property and that she relied on that misrepresentation. Of course, Wanda would fail if she lacked knowledge of the property's existence.

The actual fraud test is analytically more tenable than the constructive fraud approach.¹²⁴ The disadvantage is that unless Horace misrepresents his property interests, Wanda will be precluded from acquiring an interest in his estate. This result allows Horace to disinherit Wanda rather easily and neither balances the competing interests of the surviving spouse and children from a prior marriage nor promotes the goals of an elective share system. Next, it is necessary to examine statutory models and to analyze whether they provide a better framework for resolving Wanda's problems in the two alternatives posited: when Horace transfers his property before he contemplates marriage and when he makes the conveyances during engagement.

IV. THE STATUTORY MODELS

The vast majority of elective share statutes do not provide for a right of election against testamentary substitutes created prior to marriage or antenuptial transfers made in fraud of marital rights. These statutes are silent as to whether the common-law doctrine has been statutorily overruled or whether it co-exists with the statutory schemes. Several states have addressed, explicitly or indirectly, the issue of the includability of transfers prior to marriage in the surviving spouse's elective share. There are basically two approaches. One, exemplified by the Wisconsin and Missouri statutes, codifies the common-law rule.¹²⁵ The other establishes objective criteria to determine when a transfer will be deemed a testamentary substitute. The subjective approach embodied in the Wisconsin statute will be considered first.

A. Wisconsin

The Wisconsin elective share statute grants the surviving spouse a right of election against marital and deferred marital property owned at death as well as transfers of property that are included in the augmented marital property estate based on objective criteria.¹²⁶ In addition, section

124. See Bregy & Wilkinson, *supra* note 33, at 76.

125. Mo. ANN. STAT. § 474.150 (Vernon Supp. 1990); Wis. STAT. ANN. § 861.17(1) (West 1991).

126. Wis. STAT. ANN. §§ 861.01-07 (West 1991). Wisconsin's augmented estate concept resembles the UPC model in that it includes retained interest transfers. Like the redesigned UPC, it also "applies to life insurance, accident insurance, joint annuities, pensions and other arrangements under which property is payable to a person other than the surviving spouse or the estate of the decedent spouse." *Id.* § 861.05 comment. Pre-1990 UPC § 2-202(1) excluded such arrangements when the beneficiary was not the spouse.

861.17 of the Wisconsin code authorizes a surviving spouse to elect against “property arrangement[s] in fraud of the rights of the surviving spouse.”¹²⁷ Such arrangements include “[a]ny transfer or acquisition of property, regardless of the form or type of property rights involved, made by the decedent during marriage or *in anticipation of marriage for the primary purpose of defeating the rights of the surviving spouse.*”¹²⁸

Unlike most elective share schemes, Wisconsin’s statute, which is based on the Uniform Marital Property Act, limits the right of election to property acquired during marriage except to the extent the property was transferred in fraud of the right of election either before or during marriage. In such cases, the transferred property will be subject to the right of election whether the decedent conveyed it outright or retained an interest in it.

As originally adopted, section 861.17 provided for an exception to the fraud rule if the decedent transferred property prior to or within one year of marriage for the benefit of children from a prior marriage.¹²⁹ Historically, this exception to the common-law doctrine was recognized by a minority of jurisdictions if the children’s parents agreed that the surviving parent would bequeath property generated by their marriage to their children and the amount of property transferred to the children was not disproportionately large. The repeal of this section, however, does not mean that such transfers are necessarily fraudulent. Under section 861.17(1)(a), the court has the discretion to determine whether the “primary purpose” of the transfer was to disinherit the surviving spouse or to provide for the children. Assuming that the main purpose of the disposition was to confer a benefit on children from a prior marriage, such an arrangement may be valid.

Of course, determining the decedent’s primary purpose in making the property disposition will present the same problems under the current Wisconsin statute as existed at common law. Although the legislative committee that drafted the section recognized these difficulties and considered the objective approaches of New York and Pennsylvania, it

127. WIS. STAT. ANN. § 861.17(1) (West 1991).

128. *Id.* § 861.17(1)(a) (emphasis added). Chapter 852 of the Wisconsin statute deals with intestacy. Chapter 861 relates to the forced share.

129. Section 861.17(2) originally provided: “An arrangement made before marriage, or within one year after marriage, or prior to April 1, 1971, to provide for issue by a prior marriage is not a fraudulent property arrangement within the meaning of this section.” This subsection was repealed in 1985. The comment to this section merely indicates that “[b]ecause persons now married may have intended to provide for issue by a prior marriage, they are able to do so within a year after the effective date of this Code without danger of having such arrangements challenged as a fraud on the rights of the spouse.” *Id.* comment.

decided to recommend a codification of the common-law approach that "has the advantage of being familiar."¹³⁰

A justification based on familiarity is questionable particularly because, commencing with its frequently cited decision of *Sederlund v. Sederlund*¹³¹ in 1922, the Wisconsin Supreme Court has apparently never invalidated a prenuptial transfer of property. The actual protection af-

130. The legislature reviewed and specifically rejected the approach used in New York and Pennsylvania. STATE OF WISCONSIN SENATE BILL 5, at 162 (1969). The legislative history also indicates that the section's "interest is to fortify [the] judicial doctrine and give it procedural shape" particularly "in light of the abolition of inchoate dower by 861.03." *Id.* Thus, the section would reach "deliberate plans to deplete the probate estate in order to defeat election by the surviving spouse" regardless of whether the transfer was of real or personal property. *Id.*

A report of the Wisconsin Bar Association explained the purpose of the proposed statute:

The third preliminary draft of Sec. 6 . . . was an attempt to preserve the existing law regarding the right of the surviving spouse to reach various kinds of transfers made by a decedent during lifetime in the extreme case where the transfer was deliberately made to put the property out of reach of the spouse and any right to elect.

PROBATE CODE COMMITTEE, STATE BAR OF WISCONSIN, REAL PROPERTY, PROBATE AND TRUST LAW SECTION WORKING PAPERS, at 170 (1962-66) [hereinafter WISCONSIN BAR REPORT]. The Report criticized the decision in *In re Estate of Mayer*, 26 Wis. 2d 671, 133 N.W.2d 322 (1965), which reaffirmed that right, but indicated that the challenge must be part of an exercise of the right of election, must make the property part of the estate for all purposes, and must not merely be to reach the statutory share. *Id.* Section 6 was redrafted to avoid requiring the widow to elect against the will in order to recover a share of fraudulent transfers on the theory that the surviving spouse might not learn of the fraud until after the time to file a notice of election expired. *Id.* at 181.

In its explanation of § 6, which applies to personality and realty, the Wisconsin Bar Report noted that although the Wisconsin common-law approach has been criticized by legal scholars, the "present flexible test seems to have worked satisfactorily to reach the exceptional cases where one spouse has sought deliberately to defeat the rights of the other by transferring personality during lifetime." *Id.* at 179. The types of transactions were intentionally not enumerated. This conclusion is subject to question because, as the Report noted, "case law is extremely meager." *Id.* at 192. No Wisconsin case was found in which a surviving spouse was successful in invalidating a transfer in whole or in part as fraudulent. The burden of proving fraud in all cases is on the surviving spouse. The new section also gives the court discretion to "reduce or eliminate any share for the survivor if the decedent had reason to disinherit the surviving spouse, as where the couple have separated." WISCONSIN SENATE BILL 5, at 147 (1969).

The State Bar of Wisconsin recommended that "[t]he court's power to provide for the deserving spouse . . . should extend to any assets transferred gratuitously by the husband during lifetime . . . under circumstances which would include such assets within the taxable estate for federal estate tax purposes." WISCONSIN BAR REPORT, *supra*, at 106. This recommendation was not accepted.

131. 176 Wis. 627, 187 N.W. 750 (1922) (upholding the transfer of real estate to the children of a first marriage made before the husband met his second wife). See also WISCONSIN SENATE BILL 5, at 147 (1969).

forsaken to a surviving spouse in view of the legislative adoption of this judicial doctrine seems more theoretical than real. Furthermore, the adoption of a subjective fraud doctrine to determine the validity of antenuptial conveyances appears to contradict the objective approach¹³² adopted by the Wisconsin legislature in deciding whether post-marriage testamentary substitutes are includable in the augmented estate.¹³³

Application of the Wisconsin statute to Case 1 would result in the excludability of the transferred property from the augmented estate. Horace did not create the testamentary substitutes in Case 1 for the primary purpose of defeating Wanda's marital rights. Indeed, he was not contemplating marriage at the time of the transfer. He was motivated by a desire to benefit his children. Accordingly, Wanda would have no claim against the transferred property despite Horace's retention, during their marriage, of a life estate and a power to dispose of the transferred property until his death. If the transfers were made during marriage, the property would have been included in the augmented estate to the extent it constituted marital or deferred marital property. If, however, Horace made the transfers after his engagement to Wanda without notice to her, a court would have to resolve the difficult question of whether Horace's primary motivation was to disinherit Wanda or to provide for his children. Certainly, any argument by Horace that the substantial post-engagement transfers to his children were not made for the primary purpose of defeating Wanda's rights as a surviving spouse is disingenuous.

B. The Objective Models

The appropriate treatment of Horace's premarriage transfers will now be examined in the context of elective share statutes adopting objective criteria.

1. *Pennsylvania.*—The Pennsylvania statute formulates an objective standard for including a conveyance of property in the augmented estate. On its face, section 2203 ignores whether the transfer occurred before or during marriage, only with respect to "[p]roperty conveyed by the decedent during his lifetime to the extent that the decedent at the time of his death had a power to revoke the conveyance or to consume, invade or dispose of the principal for his own benefit."¹³⁴ The other

132. Some would argue that while an objective approach is more appropriate for marital transfers, a subjective approach is desirable for antenuptial conveyances. This rationale suggests that only those transfers made with an intent to disinherit and without notice to the intended spouse should be subject to the right of election. The author rejects this view. See *infra* text accompanying notes 256-60.

133. WIS. STAT. ANN. § 861.05 (West 1991).

134. 20 PA. CONS. STAT. ANN. § 2203(a)(3) (Purdon Supp. 1991).

categories of testamentary substitutes include retained life interests, joint tenancies, annuities, and transfers in contemplation of death that were created during marriage.¹³⁵

Why does the Pennsylvania statute seem to treat property conveyed subject to a power to revoke or consume, invade or dispose of the principal more inclusively? Neither the statute itself nor the official comment is illuminating. The comment merely indicates that section 2203 was influenced by the UPC.¹³⁶ The pre-1990 UPC, however, pointed to section 2203's predecessor, section 11 of the Pennsylvania Estates Act¹³⁷ as a model.¹³⁸ The influence of the pre-1990 UPC helps to explain why the other categories of testamentary substitutes are limited to post-marriage transfers.¹³⁹ Nevertheless, the comment to section 2203 does not expressly adopt the UPC's rationale for excluding premarital transfers from the right of election.¹⁴⁰ Furthermore, the pre-1990 UPC included transfers of property subject to a power to revoke or invade, or dispose of the principal in the augmented estate only if made during marriage.

The answer appears to lie in the fact that section 2203(a)(3) is the direct successor to section 11,¹⁴¹ while the other subsections of the

135. *Id.* § 2203(a)(2), (4)-(6).

136. *Id.* § 2203 comment.

137. 20 PA. CONS. STAT. ANN. § 2203(a)(2)-(6) was originally enacted in 1947 as § 11 of the Estates Act and was adopted in 1978 as the successor to 20 PA. STAT. ANN. § 6111 and 20 PA. STAT. ANN. § 301.11(a).

138. U.P.C. § 2-202 comment (1983).

139. U.P.C. § 2-202(1) included only testamentary substitutes created during marriage in the augmented estate. See *infra* note 218 and accompanying text.

140. See 20 PA. CONS. STAT. ANN. § 2203 comment (Purdon Supp. 1991).

141. Section 11 of the Estates Act provided in pertinent part:

A conveyance of assets by a person who retains a power of appointment by will, or a power of revocation or consumption over the principal thereof, shall at the election of his surviving spouse, be treated as a testamentary disposition so far as the surviving spouse is concerned to the extent to which the power has been reserved.

20 PA. CONS. STAT. § 301.11 (Supp. 1971). Parenthetically, § 2203(a)(3) is more limited than § 11 because it does not "include property which the decedent has given away absolutely and cannot recapture for his own benefit, even though he has retained a power of appointment which cannot be exercised in his favor during his life." 20 PA. CONS. STAT. § 2203 (Purdon Supp. 1991) (quoting Official Comment 1978). See *In re Estate of Behan*, 399 Pa. 314, 160 A.2d 209 (1960) (property subject to the decedent's special power of appointment is includable in the net estate when the decedent is also the donor of the power). Furthermore, unlike § 11, § 2203(a)(3) explicitly requires that the decedent's retained power be exercisable for the decedent's benefit. The words "power of consumption" in § 11, however, have been interpreted to mean a power to consume for the decedent's benefit. See *In re Estate of Schwartz*, 449 Pa. 112, 116, 295 A.2d 600, 603 (1972) (plurality opinion) (indicating that prior decisions under § 11 "involved situations where the interest retained by the decedent could be exercised to his own advantage"); *Longacre v. Hornblower*

Pennsylvania statute were added later, apparently in response to the UPC.¹⁴² When adopted, section 11 did not expressly exclude antenuptial transfers from the forced share on the theory that the critical factor in including transfers of property in the augmented estate is the nature of the decedent's power over the property and not the time of transfer. According to Professor Bregy, a leading commentator on and drafter of the Pennsylvania Estates Act, the legislature was advised to focus on the extent to which a decedent retained attributes of ownership over the transferred property at death and not the donor's intent in making the transfer.¹⁴³

& Weeks, 83 Pa. D. & C. 259, 262 (1952) (stating that "the legislature was using the word 'consumption' according to its common meaning, and intended to apply the act to all conveyances of assets subject to a power in the transferor to reacquire beneficially the property conveyed").

Minnesota, until recently, had a statute that appeared to have been modeled on § 11 of the Pennsylvania statute. *See* MINN. STAT. § 525.213 (1969). Without limiting its coverage to transfers during marriage, the elective share statute included transfers of property pursuant to which the decedent retained a testamentary power of appointment or a power of disposition or consumption at death. Thus, the time of the transfer and the intent with which it was made should have been irrelevant with respect to the includability of the transfer. *See Note, The Minnesota Legislature, 1969 Regular Session*, 54 MINN. L. REV. 1029, 1033 (1969) (concluding that under the Minnesota statute, the decedent's control over the property at death is determinative). Unlike § 11, which was interpreted to include joint tenancies, the Minnesota statute was held to exclude such dispositions of property. *In re Estate of O'Brien*, No. C7-87-1718 (Minn. Ct. App., Jan. 5, 1988). However, the court's reasoning appears to be questionable. First, the court stated that a joint tenancy was not a testamentary power of appointment, but did not discuss whether it was subject to a power of consumption or disposition. *Cf. O'Connell Estate*, 16 Pa. Fiduc. Rep. 491, 493 (Orphan's Ct. 1966) (holding that a joint tenancy with right of survivorship created prior to, but not in contemplation of marriage was subject to a right of election because "it was entirely within the [husband's] power . . . during his lifetime to consume his undivided one-third interest, including his right of survivorship"). Second, the court seemed to be influenced by the particular facts, rather than a proper analysis of the statutory language. In *O'Brien*, the decedent husband and the surviving spouse, although not legally separated, maintained separate residences for 15 years and made separate property arrangements. All of the husband's property was in joint tenancies with his daughters and sister with whom he had close relationships; they had frequently played the role of care-givers. In addition, the widow had sufficient independent means. Thus, in the absence of policy reasons supporting the right of election, the court found that their post-wedding joint tenancies were justifiable property dispositions.

The repeal of § 525.213 in 1985 was effective for estates of decedents dying after December 31, 1986. This section was replaced by provisions modeled on the UPC. *See* MINN. STAT. ANN. § 524.2-201 (West Supp. 1991). Thus, joint tenancies, like other testamentary substitutes, are subject to the right of election only if they are created during marriage. *Id.* § 524.2-202(1)(iii).

142. *See* 20 PA. CONS. STAT. ANN. § 2203 comment (Purdon Supp. 1991).

143. *See* B. BREGY, PENNSYLVANIA INTESTATE, WILLS AND ESTATES ACT OF 1947 5867 (1949). Bregy, who played a significant role in drafting the Estates Act, concluded

Consistent with this commentary, the lower courts in Pennsylvania interpreted section 11 to apply to all conveyances subject to retained powers, regardless of whether they were made before or during marriage. This construction was based on the statutory language that did "not limit itself to the conveyances made by a married person."¹⁴⁴ For the most part, the cases decided under section 11 ignored the fraud on marital rights decisions. Therefore, under these cases, it was immaterial whether the conveyance was made prior to the engagement or even before the couple met. Instead, the courts focused on the decedent's retention of a power of consumption over the assets transferred until death and not on any actual fraud committed against the surviving spouse.

The lower court decisions appear to have reached the correct result both in their interpretation of section 11 and in their failure to rely on the doctrine of antenuptial transfers in fraud of marital rights. As the court in *Blair Estate*¹⁴⁵ indicated, section 11 applied to "a person," not to a married person. In addition, it did not require that the transfer occur during marriage. The omission of the words "married" or "during marriage" should be material in an elective share statute that defines the parameters of a right of election.¹⁴⁶

Such an interpretation also accords with the policy behind the statute. To the extent that the goal of the legislature was to include conveyances of property in the augmented estate if the decedent retained a power to revoke the disposition or consume the principal (*i.e.*, retained economic control over the property), then the time the decedent made the disposition is immaterial. In essence, retention of this kind of control over the

that "an antenuptial transfer could be attacked under the present provision, and it probably would make no difference how long before the marriage the transfer was made." *Id.* But cf. *Smigell v. Brod*, 366 Pa. 612, 79 A.2d 411 (1951) (where the decedent conveyed real property subject to a retained life estate prior to an engagement to marry, the transferred property was not subject to the surviving spouse's share).

144. *Blair Estate*, 42 Pa. D. & C.2d 223, 228 (1967). In *Blair*, the court subjected funds included in a pension plan to the wife's right of election because the decedent contributed the funds prior to their marriage. The court held that the decedent retained a power of consumption over the contributions to the pension fund up to his death "even though it would require that he terminate his employment to do so." *Id.* at 227. Although the court did not refer to the federal estate tax cases, the same argument was used to include retirement benefits in the decedent's estate under Internal Revenue Code § 2038. See *Estate of Tully*, 528 F.2d 1401 (Ct. Cl. 1976).

145. 42 Pa. D. & C.2d 223 (1967).

146. See Mo. ANN. STAT. § 474.150(2) (Vernon Supp. 1990) ("Any conveyance of real estate made by a *married* person . . . is deemed to be in fraud of the marital rights of his spouse"); N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1) (McKinney 1981) (including as testamentary substitutes those dispositions "effected . . . after the date of the marriage"); 20 PA. STAT. ANN. § 2203(a)(2) (Purdon Supp. 1991) ("Income . . . of property conveyed by the decedent *during the marriage*" subject to right of election).

property is tantamount to property ownership; therefore, any conveyance by the decedent that involves the retention of control over the principal at the time of death and that removes property from the probate estate should be a fraud on the marital rights of the surviving spouse in the objective sense.

Because the theory behind section 11 was to treat assets that the decedent was able to dispose of or enjoy during marriage the same as property that was titled in the decedent's name, both types of conveyances should be subject to the elective right. Consequently, under section 11, testamentary substitutes were presumptively fraudulent and could be reached by the surviving spouse without requiring proof of actual fraud by the decedent. In the absence of an actual fraud requirement, the timing of the conveyance was immaterial. This fact was recognized by the lower courts, but not by the Pennsylvania Supreme Court in *Estate of Kotz*.¹⁴⁷

If not for *Estate of Kotz*, this analysis would apply with equal force to the interpretation of “[p]roperty conveyed by the decedent during his lifetime” under section 2203(a)(3). Despite the absence of language in section 11 restricting its applicability to conveyances during marriage, the Pennsylvania Supreme Court in 1979 limited the application of section 6111, the successor to section 11 and the predecessor of section 2203, to “conveyances during marriage.”¹⁴⁸ In reaching its decision, the court was influenced by the heading of section 6111: “Conveyances to Defeat Marital Rights.” The court interpreted this section to exclude premarital conveyances. This interpretation has no support in the legislative history of either section 11 or section 6111. As Bregy indicated, section 11 reached antenuptial transfers of property that otherwise met the requirements of the statute.¹⁴⁹ Furthermore, it ignored the possibility that certain prenuptial transfers are made to defeat marital rights.

Nevertheless, in view of the gloss imposed by the Pennsylvania Supreme Court's decision in *Estate of Kotz*, it is reasonable to ask whether a similar interpretation would be placed on section 2203(a)(3), which was enacted in 1978, prior to the *Kotz* decision.¹⁵⁰ If the legislature wanted to limit testamentary substitutes to those created during marriage, it would have used identical language for subsections 2203(a)(2) to

147. 486 Pa. 444, 406 A.2d 524 (1979).

148. *Id.* at 451, 406 A.2d at 531 (joint tenancy with a right of survivorship created eight years before marriage is not subject to the right of election).

149. B. BREGY, *supra* note 143, at 5867.

150. As Justice Roberts pointed out in his concurrence, § 2203(a)(4) applies to joint tenancies with a right of survivorship created by the decedent after marriage and therefore, *Kotz* would have been decided the same way under the current statute. *Kotz*, 486 Pa. at 459, 406 A.2d at 532.

2203(a)(6), particularly in view of the more expansive interpretation of section 11 by Bregy and several lower court decisions.¹⁵¹ On the other hand, the legislature has not amended the statute since *Estate of Kotz*, which may indicate its agreement with the decision.

Given the legislature's rejection of the intent test and its reliance on objective criteria to evaluate inter vivos transfers, an expansive interpretation of section 11, as opposed to the limited reading by *Estate of Kotz*, seems preferable and more consistent with the goals of Pennsylvania's elective share statute. In fact, the legislature should reconsider section 2203's inclusion of other testamentary substitutes in the augmented estate only if created during marriage. Certainly, a decedent retains no greater power over a revocable trust or property transferred subject to a retained power to invade the principal than over a joint tenancy with a right of survivorship. In addition, the beneficial interest is not necessarily greater when the decedent retains a power to invade the corpus for his or her benefit rather than a life estate. In each case, the focus should only be on the extent of the decedent's power over or interest in the transferred property at death and not on whether the transfer preceded or followed marriage.¹⁵² This rationale applies with equal force to gifts in contemplation of death.¹⁵³ Although the decedent's legitimate estate and tax planning objectives should be safeguarded, an individual should not be able to deplete the augmented estate by making large antenuptial gifts even if death follows within a year of marriage.¹⁵⁴

151. B. BREGY, *supra* note 143, at 5867. See *O'Connell Estate*, 16 Pa. Fiduc. Rep. 491, 492-93 (1966) (where joint tenancy with a right of survivorship was created before the marriage was contemplated and was not for the purpose of defrauding the widow, the surviving spouse could elect against the property under § 11 because the decedent had a lifetime power to *consume* his interest); Clephane Trust, 36 Pa. D. & C.2d 386 (1965) (allowing a right of election against a revocable trust created more than two years before the marriage and without intent to defraud later wife); *Hagy Estate*, 15 Pa. Fiduc. Rep. 456, 461 (1964).

152. The only relevance that a premarriage transfer should have is if policy dictates that the surviving spouse receive a lesser fraction of the augmented estate; in that event, the fraction should relate to the length of the marriage and not the time the testamentary substitute was created. Cf. Langbein & Waggoner, *supra* note 1 (advocating adjustments to the forced share based on duration of marriage). Another relevant factor might be whether the decedent had children from a prior marriage.

153. 20 PA. CONS. STAT. ANN. § 2203(a)(6) (Purdon Supp. 1991).

154. Because § 2203(a)(6) only captures gifts made within one year of death and then only if "the aggregate amount so conveyed to each donee exceeds \$3,000," if it applied to antenuptial transfers, the marriage could only have lasted less than one year. *Id.* In cases of such a short marriage, some have argued that it is inequitable to give the surviving spouse the same portion of the decedent's estate as the survivor would have been entitled to had the marriage lasted 30 years. See, e.g., Langbein & Waggoner, *supra* note 1, at 314-21 & n.8. This objection can best be dealt with, if it should be at all, by adjusting the fractional share rather than excluding premarriage transfers. See U.P.C. § 2-201(a) (Supp. 1991).

The other rationale for excluding antenuptial transfers from the augmented estate, that is, preserving dispositions of property for the benefit of children from a prior marriage,¹⁵⁵ does not support the distinction made by the Pennsylvania statute if the limitation in *Estate of Kotz* does not apply to section 2203(a)(3). No logical reason exists to suppose that a parent will be less likely to use a retained power to benefit a child rather than a life estate or joint tenancy.¹⁵⁶

More troubling is the silence in the legislative history of section 11 and its successors concerning the interplay, if any, between the elective share statute and the transfer in fraud of marital rights doctrine.¹⁵⁷ Assuming that section 2203(a)(3) applies to prenuptial transfers, conveyances subject to retained powers of revocation or invasion for the decedent's benefit should no longer be subject to the doctrine because this section makes the existence of fraud irrelevant to the includability of the transferred property in the augmented estate. Conveyances covered by section 2203(a)(2), (4), (5), and (6) would still be subject to the common-law doctrine.¹⁵⁸ Of course, the doctrine presumably applies to outright conveyances in which the decedent retained no interests or powers exercisable for his or her own benefit. These transfers are not testamentary substitutes even if created during marriage.

The cases decided subsequent to the adoption of section 11 involving the antenuptial transfer in fraud of marital rights doctrine shed no light

155. See, e.g., U.P.C. § 2-202 comment (1983). But see Wis. STAT. ANN. § 861.17(2) (repealed 1985). There is, however, no evidence that in Pennsylvania, antenuptial transfers in fraud of marriage for the benefit of a child were not subject to attack. This exception was originally enunciated in the English cases. See *supra* note 80.

156. This author's review indicates, however, that a majority of cases dealing with transfers in fraud of marriage involved retained life interests. This should not be determinative because most of the cases dealt with transfers of land in fraud of dower where the most common type of retained interest would probably be a life estate or a joint tenancy.

157. See, e.g., O'Connell Estate, 16 Pa. Fiduc. Rep. 491 (Orphan's Ct. 1966). The court found that a joint tenancy with right of survivorship held by the decedent with his sisters was subject to the widow's right of election pursuant to § 11 despite its creation prior to marriage and merely noted that "[t]here is no claim that the deed was made for the purpose of defrauding the rights of the widow . . . or that the decedent had any intention of marriage at the time the deed was executed." *Id.* at 492-93. The court seemed to assume that because the deed was not executed in contemplation of marriage, it was valid. The court neglected to indicate, however, that as the doctrine was applied in Pennsylvania, it only invalidated a transfer to the extent of the surviving spouse's rights in the property. Thus, assuming the requirements of the doctrine of antenuptial transfers in fraud of marital rights and § 11 were met, the results should not have been different.

158. Commencing with *Kirk v. Kirk*, the Pennsylvania Supreme Court, in imposing an actual fraud requirement, took a more stringent view of the requirements necessary to invalidate a transfer as a fraud on marital rights. For a discussion of the evolution of this doctrine in Pennsylvania, see Bregy & Wilkinson, *supra* note 33.

on the relationship between this rule and the elective share statute.¹⁵⁹ They only demonstrate that the doctrine continues to be applied without regard to the elective share statute. The surviving spouse, however, will have greater difficulty proving fraud now than in the period preceding the Pennsylvania Supreme Court's decision in *Kirk v. Kirk*,¹⁶⁰ which imposed an actual fraud test.

2. *New York*.—A different approach is followed by New York's Estates, Powers & Trusts Law (EPTL),¹⁶¹ which explicitly includes in the net estate only those testamentary substitutes created during marriage.¹⁶² Like the Pennsylvania statute and the UPC,¹⁶³ the EPTL adopts objective criteria to determine the includability of inter vivos transfers in the elective share, but does not limit transfers subject to retained powers to dispositions for the decedent's benefit.¹⁶⁴ The EPTL is silent as to the treatment of antenuptial transfers made in fraud of marital rights.

The legislative history of the EPTL is not helpful in determining whether the common-law doctrine of antenuptial transfers in fraud of marital rights survives the enactment of the objective standards adopted in section 5-1.1(b)(1). A legislative report of the Temporary State Commission merely notes that section 18-a (enacted as EPTL section 5-1.1)

159. See, e.g., *Brinko v. Redden*, 402 Pa. 408, 167 A.2d 467 (1961) (where a decedent transferred real property apparently reserving a life estate during his first marriage and two years before his second marriage, the conveyance was not in fraud of the second wife's marital rights); *Smigell v. Brod*, 366 Pa. 612, 79 A.2d 411 (1951) (enunciating the rule that when a husband transferred property subject to a retained life estate prior to engagement, the surviving spouse could not acquire an equitable interest in the property and therefore, could not set aside the conveyance as fraudulent); *Durant v. Durant*, 294 Pa. Super. 202, 439 A.2d 821 (1982) (upholding the decedent's transfer of a tavern to his brother six months prior to marriage because the wife could not sustain the burden of proof that she knew of and relied on her husband's possession of property, the husband received \$10,000, and the transfer may have occurred prior to the engagement). Cf. *Andrikanics v. Andrikanics*, 371 Pa. 222, 223, 89 A.2d 792, 793 (1952) (where a wife challenged her husband's transfer of property to the children of his first marriage two weeks before the marriage on the ground of undue influence, the court, in upholding the transfer, stated that "[a] question might be raised as to whether . . . the wife-plaintiff has the status in this proceeding to demand relief inasmuch as she was not married to the husband-plaintiff at the time the conveyance in question was made"). The *Andrikanics* court's reliance on *Smigell* appears to be misplaced because *Smigell*, like *Brinko*, involved a transfer prior to the engagement.

160. 340 Pa. 203, 16 A.2d 47 (1940). It should be noted that *Kirk v. Kirk* was decided seven years before the adoption of § 11. For further discussion of this case, see *supra* note 112.

161. N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1) (McKinney 1981).

162. *Id.*

163. U.P.C. § 2-202(b)(2) (Supp. 1991).

164. N.Y. EST. POWERS & TRUST LAW § 5-1.1(b)(1) (McKinney 1981).

applies to “[o]nly such transactions effected by the decedent after the marriage and after the effective date of the bill.”¹⁶⁵ The Commission enumerated “areas of related subjects which are not covered herein but which could be profitably studied”¹⁶⁶ and asked whether “the present law with respect to transfers made between the engagement and the marriage [should] be changed.”¹⁶⁷ This question was never answered and it leaves open the possibility that a subjective test incorporating the common-law doctrine may be applied to testamentary substitutes created during the engagement period. The vehement rejection by the Bennett Commission, which drafted section 5-1.1 and the legislative reports, of applying a subjective test to testamentary substitutes created during marriage casts doubt on the doctrine’s continued viability.¹⁶⁸ This is particularly true with respect to outright transfers.¹⁶⁹ In addition, the report’s query assumes that transfers made prior to an engagement will be excluded from the right of election. Most troublesome is the implicit assumption that testamentary substitutes created prior to marriage should be regarded differently for elective share purposes than assets that are acquired before marriage and titled in the decedent’s name at death. The reports accompanying the EPTL give no rationale for this distinction.¹⁷⁰

165. THIRD REPORT, *supra* note 54, at 24.

166. *Id.* at 118.

167. *Id.* at 118-19 n.14(e) (citing W. MACDONALD, *supra* note 33, at 354 app. C); Vaughan, *supra* note 33, at 178.

168. In a note to the proposed statute, the Report indicates that the “act . . . is intended to provide a comprehensive statement as to the rights of a surviving spouse with respect to inter vivos transactions made by the decedent, as well as property disposed of by will.” THIRD REPORT, *supra* note 54, at 116.

169. *Id.* at 138 (“It is submitted that to protect freedom of alienation of property every outright and absolute transaction should be immune from the attack of the surviving spouse even if it is gratuitous.”).

170. In 1985, a proposed amendment to EPTL § 5-1.1(b)(1) included testamentary substitutes created prior to marriage in the net estate. The report, in support of the proposal introduced as Senate Bill No. 5058, explained that the appropriate test for inclusion as a testamentary substitute is revocability, not whether it is created before or during marriage. See LEGISLATURE OF NEW YORK, MEMORANDUM OF THE LAW REVISION COMMISSION RELATING TO THE RIGHT OF ELECTION OF A SURVIVING SPOUSE, LEGIS. DOC. NO. 65[D], at 2 (1985); LEGISLATURE OF NEW YORK, RECOMMENDATION OF THE LAW REVISION COMMISSION TO THE 1985 LEGISLATURE RELATING TO THE RIGHT OF ELECTION OF A SURVIVING SPOUSE, at 3 (1985). This amendment was never adopted by the legislature. Last year, a proposal and memorandum in support of the proposal submitted by the Trusts and Estates Section of the New York State Bar Association recommended changes in EPTL § 5-1.1 that generally broadened the rights of the surviving spouse. See N.Y. BAR REPORT, *supra* note 1. These changes, however, did not include subjecting premarriage testamentary substitutes to the right of election. Most recently, the Advisory Committee Report made broad recommendations to revise both the New York elective share and

The few cases involving premarriage transfers decided after the enactment of EPTL section 5-1.1¹⁷¹ shed no light on the continued viability of the doctrine of antenuptial transfers in fraud of marital rights.¹⁷² *In re Scheiner*¹⁷³ is a recent example. In *Scheiner*, the decedent, prior to marriage, purchased Treasury bills registered in the names of the decedent and his sister, payable to the survivor. The court held that the surviving spouse could not elect against the Treasury bills, even though they rolled over during marriage, because the decedent purchased them before marriage.¹⁷⁴ The court did not discuss whether the decedent purchased them in contemplation of marriage, but applied a mechanical test to determine whether the inter vivos transfers met the requirements of a testamentary substitute.¹⁷⁵

The paucity of cases in the post-EPTL section 5-1.1 era is surprising in light of the comparatively large number of cases that applied the doctrine to premarriage transfers made before the effective date of the act. This is particularly striking in light of New York's liberal interpretation of the doctrine covering premarriage transfers of personality as well as real estate. To the extent that section 5-1.1 restricts a spouse's

intestacy statutes. Among its recommendations regarding the elective share was a proposal to include as testamentary substitutes: (a) insurance policies on the decedent's life if the decedent retained any incidents of ownership or transferred them within one year of death; (b) pension, profit sharing, and deferred compensation plans and IRAs; (c) United States government bonds payable on the decedent's death to a third party; (d) dispositions made during marriage subject to retained life interests; (e) revocable transfers; (f) general lifetime powers of appointment; and (g) certain inter vivos transfers made within one year of death that exceed the gift tax exclusion under I.R.C. § 2503(b) and (e) to the extent that the "decedent did not receive full and adequate consideration in money or money's worth for such transfers." ADVISORY COMMITTEE REPORT, *supra* note 1, at 15-16 & app. C. The Report rejected modeling the elective share statute on either the community property or the equitable distribution models. *Id.* at 8-9.

171. EPTL § 5-1.1 became effective in 1967 and applied to testamentary substitutes created on or after September 1, 1966 if the decedent and the surviving spouse were married prior to their establishment and the decedent executed a will or died intestate on or after that date. N.Y. EST. POWERS & TRUSTS LAW § 5-1.1 app. (McKinney 1981).

172. At least one commentator appears to assume that the doctrine is still valid. See Note, *EPTL § 5-1.1; Buy and Sell Agreement Directing Payment to Third Party Upon Death of Testator Held a Testamentary Substitute*, 58 ST. JOHN'S L. REV. 219, 223 n.166 (1983) (indicating that a transfer of realty while engaged would be void to the extent of a widow's dower rights).

173. 141 Misc. 2d 1037, 535 N.Y.S.2d 920 (1988). For a commentary on *Scheiner*, see Gibbs & Ordover, *Right of Election by Surviving Spouse*, N.Y.L.J., Sept. 13, 1989, at 3, col. 4.

174. *Scheiner*, 141 Misc. 2d at 1038, 535 N.Y.S.2d at 921. The surrogate determined that Treasury bills could never be testamentary substitutes because of the exemption in EPTL § 5-1.1(b)(2)(C) excluding "United States savings bonds payable to a designated person." *Id.*

175. Cf. Estate of Walton, 56 A.D.2d 436, 392 N.Y.S.2d 621 (1977).

ability to challenge prenuptial conveyances of property, this result is undesirable.

*LeStrange v. LeStrange*¹⁷⁶ crystallized the inherent dangers to a spouse in a multiple marriage society in which the surviving spouse's interest may be completely sacrificed to the legitimate interests of children from a prior marriage.¹⁷⁷ *LeStrange* involved a husband and wife who married late in life and had children from prior marriages. After their engagement, Mr. LeStrange represented to his fiancee that he owned real property and a substantial sum of money. At the urging of his children, and without informing his prospective wife, he transferred all of his property to a trust approximately six weeks before his marriage. The trust directed Mr. LeStrange, as trustee, to pay himself the income for life with the remainder to be divided among his children. The trustee had the power to withdraw the principal only with the consent of his sons.¹⁷⁸ The court stated that the "transfers were made in contemplation of the marriage

176. 242 A.D. 74, 273 N.Y.S. 21 (1934). Prior to the *LeStrange* decision, the doctrine was apparently only applied to antenuptial conveyances of real property.

177. Protecting the interests of children in a multiple marriage society is a central concern of the redesigned elective share provisions of the UPC. See U.P.C. art. II, pt. 2 general comment (Supp. 1991); Waggoner, *Spousal Probate Rights*, *supra* note 1. Securing the children's interest is accomplished at the surviving spouse's expense when accommodation of both parties' interests is possible.

178. Assuming that this trust was created during marriage and after the effective date of EPTL § 5-1.1, it arguably would not have been considered a testamentary substitute. Reservation of a life estate under the EPTL (unlike the Pennsylvania statute or the UPC) does not render either the life estate or the underlying transfer includable in the net estate. See N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1) (McKinney 1981). But see *supra* note 170 (recommendation of the Advisory Committee Report). Generally, a power granted by the express provisions of the instrument to invade the principal held by the decedent at death in conjunction with another person results in the includability of the principal to the extent of the power in the net estate. In this situation, the husband held that power with the beneficiaries. See *LeStrange*, 242 A.D. at 75, 273 N.Y.S. at 23. In such a case, the trust was terminable and therefore invadable as a matter of law (under EPTL § 7-1.9(a) (McKinney 1967)), not by virtue of the "express provisions of the disposing instrument." N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1)(E). The Report makes clear that a power of the decedent to terminate in conjunction with the beneficiaries is not a power within the elective share provision. See THIRD REPORT, *supra* note 54, at 130; Kwestel & Seplowitz, *supra* note 1, at 18. But see *In re Estate of Riefberg*, 58 N.Y.2d 134, 446 N.E.2d 424, 459 N.Y.S.2d 739 (1983) (finding a power to revoke under § 5-1.1(b)(1)(E) where a contract provided that it could be terminated upon the mutual agreement of the parties). A conclusion that the *LeStrange* trust would not be a testamentary substitute even if created during marriage would point to a serious deficiency in the New York statute. This deficiency could be remedied easily by amending the statute to provide that a transfer with a retained life estate constitutes a testamentary substitute or by eliminating the distinction between a power to revoke or terminate retained by the express provisions of the disposing instrument and such power arising by operation of law. See Kwestel & Seplowitz, *supra* note 1, at 24-27, 60.

and with intent to defraud the plaintiff and deprive her of her prospective rights as a surviving spouse"¹⁷⁹ without notice to her and held the transfers void under the following rule:

A man cannot deliberately, by his own purpose or through the inducement of other interested persons, strip himself of all inheritable property in fraud of his prospective wife, to whom he has represented that he has sufficient property for their comfortable support and maintenance and in which she will be entitled to share upon his decease.¹⁸⁰

Although the right to support and maintenance exists during marriage, the court explicitly recognized an inheritance right which is a mere expectancy until death.¹⁸¹ Thus, to reach an equitable result, the court essentially held that Mr. LeStrange defrauded his prospective wife of an expectancy that would not ripen into a right until death. Application of this rule, however, does not produce the most desirable result. If Mr. LeStrange had created the trust prior to his engagement (the situation in Case 1), the transfer would not have implicated the doctrine and would have been valid. At her husband's death, Mrs. LeStrange would be left with virtually no interest in his estate, although the children's interests would be secure.¹⁸² From a support and marital-sharing viewpoint, this does not promote the goals of an elective share system. On the other hand, giving the surviving spouse a right to elect against the trust property by virtue of the reserved life estate and power to invade the principal would recognize her support needs and her contribution to the marriage. This would decrease, but certainly not eliminate, the children's shares. In short, incorporating the doctrine of antenuptial transfers in fraud of marital rights into the New York elective share statute would not produce the most favorable result.¹⁸³

3. *The Uniform Probate Code.*

a. *The marital partnership theory*

The recent revisions of the UPC's elective share provisions embody a view of marriage and spousal rights at death that is dramatically

179. *LeStrange*, 242 A.D. at 76, 273 N.Y.S. at 24.

180. *Id.* In *LeStrange*, unlike many of the constructive fraud cases, the decedent engaged in actual misrepresentation with ensuing reliance by his prospective wife.

181. *Lestrangle v. Lestrangle*, 242 A.D. 74, 76, 273 N.Y.S. 23, 24 (1934). See also *Rubin v. Myrub Realty Co.*, 244 A.D. 541, 279 N.Y.S. 867 (1935).

182. This assumes that, as an older couple, significant assets were not accumulated during marriage.

183. This is independent of all other problems associated with the doctrine.

different from its predecessor. Redesigned Article II explicitly adopts the “contemporary view of marriage . . . [as] an economic partnership.”¹⁸⁴ According to Professor Lawrence Waggoner, who was primarily responsible for the reformulation of Article II, two rationales support this view. As the UPC indicates, the sharing theory derives from “an unspoken marital bargain under which the partners agree that each is to enjoy a half interest in the fruits of the marriage, *i.e.*, in the property nominally acquired by and titled in the sole name of either partner during the marriage (other than property acquired by gift or inheritance).”¹⁸⁵ Pursuant to this theory, property produced before the creation of the partnership or acquired gratuitously, and therefore not generated with the other partner’s support, is not part of the bargain and can be disposed of freely. A decedent who disinherits a surviving spouse by leaving marital property to third parties breaches the implied agreement between the spouses, and the decedent’s estate is subject to the survivor’s elective right.¹⁸⁶ A second theory incorporates a restitution concept. This theory recognizes the surviving spouse’s contribution to the marriage, whether pecuniary or nonpecuniary, and compensates the survivor for lost opportunities.¹⁸⁷

In a community property system, a marital-sharing theory results in a fifty percent spousal ownership of the marital property produced at any time during marriage. The UPC rejected this approach for administrative reasons and instead, implemented the accrual system proposed by Professors Langbein and Waggoner.¹⁸⁸ This system bases the elective share on the length of the marriage and assumes that the percentage of marital property increases with the length of the marriage.¹⁸⁹ The surviving spouse of a short-term marriage of one to two years receives three percent of the decedent’s property titled in the decedent’s name,

184. U.P.C. § 2-202 art. II, pt. 2 general comment (Supp. 1991); Waggoner, *Spousal Probate Rights*, *supra* note 1, at 349-51.

185. U.P.C. art. II, pt. 2 general comment (Supp. 1991). See M. GLENDON, *supra* note 17, at 131. This, of course, assumes that the parties did not enter into an antenuptial agreement.

186. Waggoner, *Spousal Probate Rights*, *supra* note 1, at 349.

187. *Id.* at 349-350. See M. GLENDON, *supra* note 17, at 131.

188. See Langbein & Waggoner, *supra* note 1, at 314-21.

189. The accrual system can be problematic. For instance, a surviving spouse who has been married for less than a year, and who buys a lottery ticket with marital assets, wins, puts the money in his or her own name and then dies, will be disadvantaged. The Advisory Committee Report specifically rejected the accrual system. It found this feature “undesirable” because it “essentially valued spousal rights by reference only to the length of a marriage, ignoring both the source of the marital assets and the fact that even in marriages of short duration a surviving spouse may have contributed substantially to the acquisition of family assets.” ADVISORY COMMITTEE REPORT, *supra* note 1, at 9.

augmented by certain testamentary substitutes. This share increases with each year of marriage to a maximum of fifty percent of the property.¹⁹⁰ By the end of fifteen years, virtually all of the decedent's property, the drafters of the UPC posit, will be generated by partnership efforts.¹⁹¹ At the heart of the UPC's accrual system is the premise that only property acquired by the marital partnership, absent some need of the surviving spouse that would implicate the support mechanism,¹⁹² should be includable in the augmented estate.¹⁹³ Under this theory, property acquired prior to marriage should not be subject to the right of election.

This approach reflects the concern for ease of administration and predictability in probate matters. The accrual method seeks to avoid the difficult problems frequently associated with equitable distribution and community property systems.¹⁹⁴ For example, the accrual system eliminates the need to trace the source of the assets to determine whether they were acquired before or during the marriage or by gift or inheritance.¹⁹⁵ Thus, the difficulties encountered in a community property jurisdiction or under an equitable distribution system differentiating between marital and separate property are dispensed with "by applying an ever-increasing percentage to the couple's combined assets without regard to when or how those assets were acquired, rather than applying a constant percentage (50%) to an ever-growing accumulation of assets."¹⁹⁶ Another rationale for the accrual system, although not artic-

190. U.P.C. § 2-201(a) (Supp. 1991).

191. *Id.*

192. *Id.* § 2-201(b).

193. *Id.* art. II, pt. 2 general comment.

194. See Kwestel & Seplowitz, *supra* note 1, at 8 n.23; Waggoner, *Spousal Probate Rights*, *supra* note 1, at 355-58. Professor Waggoner also notes that because a goal of the redesigned UPC is to achieve uniform state probate laws, the UPC could not incorporate the equitable distribution system that varies among the states. See J. OLDHAM, DIVORCE, SEPARATION AND THE DISTRIBUTION OF PROPERTY (1989). In its rejection of an equitable distribution model, the Advisory Committee Report stated that the use of discretionary factors and the difficulty of tracing assets are undesirable. ADVISORY COMMITTEE REPORT, *supra* note 1, at 8-9.

195. U.P.C. art. II, pt. 2 general comment (Supp. 1991).

196. *Id.* The comment explains that the elective share percentage of the couple's combined property is equated with 50% of their marital assets. It gives an example of a marriage lasting between 10 and 11 years. Under § 2-201(a), the elective share percentage is 30% and under § 2-207(a)(4), 30% of the couple's combined assets are equated with 50% of the assets acquired during the marriage (other than by gift or inheritance). Section 2-207 specifies the procedure for satisfaction of the elective share and subsection (a)(4) provides:

(a) In a proceeding for an elective share, the following are applied first to satisfy the elective-share amount and to reduce or eliminate any contributions due from the decedent's probate estate and recipients of the decedent's reclaimable estate: . . .

ulated, is that the commitment to the marriage arguably increases with its length. A surviving spouse of a longer marriage should receive a greater percentage of the decedent's estate. Basing the fractional share on a commitment theory seems to accord with the justifiable expectations of the surviving spouse.

The augmented estate against which the elective share percentage is applied also includes the value of the property owned by the surviving spouse at the decedent's death and the value of the surviving spouse's reclaimable estate.¹⁹⁷ The purpose of combining both spouses' estates is to "implement a partnership or marital-sharing theory"¹⁹⁸ as well as to ensure that the surviving spouse does not receive a windfall. Under the pre-1990 UPC, the surviving spouse's property was included in the augmented estate, but only to the extent that it was titled in the survivor's name or in the form of testamentary substitutes and was derived from the decedent.¹⁹⁹ Although the comment to redesigned section 2-202 does not specify the rationale for the prior version,²⁰⁰ a reading of the pre-1990 section and its comment indicates that the policy was to include property given to the surviving spouse by the decedent during the latter's

(4) Amounts included in the augmented estate under Section 2-202(b)(4) up to the applicable percentage thereof. For the purposes of this subsection, the "applicable percentage" is twice the elective-share percentage set forth in the schedule in Section 2-201(a) appropriate to the length of time the spouse and the decedent were married to each other.

Id. § 2-207(a)(4).

197. *Id.* § 2-202(b)(4). This subsection includes:

The value of property owned by the surviving spouse at the decedent's death, reduced by enforceable claims against that property or that spouse, plus the value of amounts that would have been includable in the surviving spouse's reclaimable estate had the spouse predeceased the decedent. But amounts that would have been includable in the surviving spouse's reclaimable estate under subsection (b)(2)(iii) are not valued as if he were deceased.

Subsection (b)(2)(iii) refers to the proceeds of insurance and retirement benefits other than Social Security.

198. *Id.* § 2-202 comment. The Advisory Committee Report rejected the UPC's distinction between assets acquired before marriage and assets acquired during marriage and its inclusion of the surviving spouse's estate in the augmented estate. The Report noted, however, that some members of the Committee were troubled by its position. ADVISORY COMMITTEE REPORT, *supra* note 1, at 8.

199. U.P.C. § 2-202(2) (1983). Of course, the consequence of including lifetime transfers to the surviving spouse is a reduction of the spouse's right of election because these transfers are credited against the elective share.

200. The comment to the redesigned UPC notes that pre-1990 § 2-202(2) included those assets of the surviving spouse derived from the surviving spouse "[u]nder a different rationale, no longer appropriate under the redesigned system." U.P.C. § 2-202 comment (Supp. 1991).

lifetime in the augmented estate.²⁰¹ Thus, the focus of the pre-1990 UPC was on the decedent's property. Its goal was not the sharing of marital property at death, which can be effected by pooling the property of both spouses regardless of its source, but rather on preventing the survivor from realizing a windfall at the decedent's death by considering the extent to which the decedent made provisions for the survivor during life. In addition to implementing the marital partnership theory, this feature of redesigned section 2-202 eliminates the need to trace the source of the survivor's property which can be an extremely arduous task, particularly in the case of a long marriage.²⁰²

b. Testamentary substitutes includible regardless of when created

Redesigned section 2-202 manifests an enhanced concern with preventing spousal disinheritance by broadening the surviving spouse's right of election against nonprobate transfers. For the first time, it extends the right of election to certain nonprobate property transferred prior to marriage or property over which the decedent acquired a power prior to marriage and "owned in substance as well as in form."²⁰³ Three categories of property fall within this group: (1) property subject to a presently exercisable general power of appointment held by the decedent; (2) property subject to the decedent's unilaterally severable right of

201. U.P.C. § 2-202(2) & comment (1983). The pre-1990 UPC included transferred property in the augmented estate when the beneficiary was the surviving spouse. Property derived from the decedent, whether from inter vivos gifts or insurance benefits paid at death, was credited against the surviving spouse's right of election under pre-1990 UPC § 2-202(2) regardless of "whether the transfers [were] made before or after marriage." *Id.* Although not explicitly articulated, the reason for including transfers prior to marriage in the augmented estate when they benefit the surviving spouse appears to have been the section's policy of protecting the surviving spouse while preventing the receipt of a windfall. In all likelihood, most of these transfers to the surviving spouse probably occurred after marriage; therefore, the distinction, from a practical viewpoint, may not have been significant. The comment explained:

[T]hus if a husband has purchased a home in the wife's name and made systematic gifts to the wife over many years, the home and accumulated wealth she owns at his death as a result of such gifts ought to, and under this section do, reduce her share of the augmented estate. . . . [I]t is obvious that this section will operate in the long run to decrease substantially the number of elections. This is . . . because the spouse can no longer elect in cases where substantial provision is made by joint tenancy, life insurance, lifetime gifts, living trusts set up by the decedent, and the other numerous nonprobate arrangements by which wealth is today transferred.

Id. § 2-202 comment.

202. This advantage is explicitly recognized in the comment. U.P.C. § 2-202 comment (Supp. 1991).

203. *Id.*

survivorship; and (3) proceeds of insurance policies in which the decedent retained certain incidents of ownership.²⁰⁴

In the first category, property is includable in the reclaimable estate as long as the decedent held the power alone immediately before death or if not, released the power or exercised it in favor of anyone other than the decedent or the decedent's estate, spouse, or surviving spouse while married to the surviving spouse and within two years of death.²⁰⁵ The property is includable regardless of whether the decedent was also the donor of the power and whether the power was created before or during marriage.²⁰⁶ The comment explains the reason for this departure from the prior Code:

The general theory of revised Section 2-202(b)(2)(i) is that a decedent who, during life, alone had a power to make himself or herself the full technical owner of property was in substance the owner of that property for purposes of the elective share. Whether the decedent created that power or it was created by another, and whether that power was created before or after the marriage, are irrelevant. The only relevant criteria are whether the decedent held that power at (or immediately before) death or (while married and during the two-year period prior to the decedent's death) irrevocably exercised or released it.²⁰⁷

Unlike the time at which the general power of appointment is created, the time of exercise or release is important for purposes of the elective

204. *Id.* § 2-202(b)(2)(i)-(iii).

205. *Id.* § 2-202(b)(2)(i). This subsection includes in the reclaimable estate: [P]roperty to the extent the passing of the principal thereof to or for the benefit of any person, other than the decedent's surviving spouse, was subject to a presently exercisable general power of appointment held by the decedent alone, if the decedent held that power immediately before his death or if and to the extent the decedent, while married to his surviving spouse and during the two-year period next preceding the decedent's death, released that power or exercised that power in favor of any person other than the decedent or the decedent's estate, spouse, or surviving spouse.

Interestingly, property subject to a power of appointment exercisable in favor of creditors of the decedent's estate would not be includable in the reclaimable estate under this paragraph. *See I.R.C. § 2041(b)(1)* (1986).

206. The pre-1990 UPC did not explicitly mention powers of appointment. Under former section 2-202(1)(ii), property subject to the decedent's presently exercisable general power of appointment would presumably have been includable in the augmented estate if the decedent created the power while married to the surviving spouse. U.P.C. § 2-202(1)(ii) (1983).

207. U.P.C. § 2-202 comment (Supp. 1991). The stated theory for including property subject to powers of appointment created prior to marriage should apply with equal force to any testamentary substitute and, in particular, to revocable trusts. *See infra* text following note 238.

share. When the decedent exercises or releases a general power of appointment, the property subject to the power will be includable in the reclaimable estate only if the decedent was married to the surviving spouse at the time of the exercise or release. This distinction makes sense for an elective share statute. If outright transfers of property are subject to the right of election,²⁰⁸ the motive for those transfers should be disinheritance. Inclusion of such property based upon this theory does not, however, mandate the use of a subjective approach. The UPC accomplishes this result by an objective scheme that includes only outright transfers made during marriage and within two years of death. This section makes the justifiable assumption that a married decedent who exercises or releases a general power of appointment (or who transfers an insurance policy or other property) in contemplation of death does so to ensure that the surviving spouse will not receive the property at death. The UPC adopts a blanket two year rule, thereby avoiding unnecessary and difficult litigation regarding the decedent's motive in making the transfer and whether it was made in contemplation of death.²⁰⁹ Essentially, the UPC adopts a *per se* rule throughout section 2-202 that any transfer or release of a power or beneficial interest in a substantial gift of property made while married and within two years of death is in fraud of the augmented estate rights of the surviving spouse and includable in the reclaimable estate.

The revised UPC takes a similar approach toward the decedent's unilaterally severable interest in property held by the decedent and someone other than a spouse with a right of survivorship.²¹⁰ As long as the decedent held an interest in the property with a right of survivorship immediately before his or her death, the decedent's interest is includable in the reclaimable estate regardless of whether the right of survivorship was created prior to marriage. Inclusion of this interest conforms to the

208. As the comment recognizes, a release of a power of appointment (or allowing it to lapse at death) is in effect a transfer of "the property subject to the power to the persons who benefit from a nonexercise of the power." U.P.C. § 202 comment (Supp. 1991). *See also* I.R.C. § 2041 (1986).

209. This approach is consistent with the prior section that included in the augmented estate certain outright transfers made within two years of death. *See* U.P.C. § 2-202(1)(iv) (1983).

210. U.P.C. § 2-202(b)(2)(ii) (Supp. 1991). This subsection includes in the reclaimable estate:

[P]roperty, to the extent of the decedent's unilaterally severable interest therein, held by the decedent and any other person, except the decedent's surviving spouse, with right of survivorship, if the decedent held that interest immediately before his death or if and to the extent the decedent, while married to his surviving spouse and during the two-year period preceding the decedent's death, transferred that interest to any person other than the decedent's surviving spouse.

policy of the elective share because the decedent had unfettered access to the property during marriage. If the decedent transferred an interest to another person within two years of death, as with the release or exercise of a general power of appointment, the property will be includable only if the decedent was married at the time of the transfer.²¹¹

In a dramatic change from the original version, the reclaimable estate of redesigned section 2-202 includes insurance proceeds payable to a person other than the surviving spouse to the extent that the decedent retained certain incidents of ownership.²¹² The exclusion of such proceeds from the augmented estate is frequently criticized on the ground that it provides an easy vehicle to disinherit a surviving spouse completely or in part.²¹³ Under the current version, the right of election extends to insurance proceeds if the decedent owned the policy, held one of the prohibited powers immediately before death, or transferred the policy to a person other than the decedent's spouse while married and within the two year period prior to death.²¹⁴ Thus, if the decedent owned the policy or held a prohibited power at death, then it is immaterial whether the decedent acquired the policy or any of the powers before or during marriage. If the decedent transferred the policy within two years of death, the proceeds will be subject to the surviving spouse's right of election only if the decedent was married to the surviving spouse at the time of transfer.

Inclusion of insurance proceeds, regardless of when the policy was acquired, represents a positive change in the UPC. The comment ac-

211. See *supra* note 208.

212. Compare U.P.C. § 2-202(1) (1983) with U.P.C. § 2-202(b)(2)(iii) (Supp. 1991). Under the latter subsection, the reclaimable estate includes:

[P]roceeds of insurance, including accidental death benefits, on the life of the decedent payable to any person other than the decedent's surviving spouse, if the decedent owned the insurance policy, had the power to change the beneficiary of the insurance policy, or the insurance policy was subject to a presently exercisable general power of appointment held by the decedent alone immediately before his death or if and to the extent the decedent, while married to his surviving spouse and during the two-year period next preceding the decedent's death, transferred that policy to any person other than the decedent's surviving spouse.

Retention of the incidents of ownership with respect to an insurance policy would render the proceeds includable in the gross estate for federal estate tax purposes under Internal Revenue Code § 2042(2). Macdonald advocates a more inclusive policy toward insurance, contending that "the surviving spouse should have a stronger claim against the husband's life insurance than she would under the present estate tax provisions." W. MACDONALD, *supra* note 33, at 278. His position imposes an undue restraint on alienation and may interfere with legitimate estate planning when the decedent has relinquished control over the property. See Kwestel & Seplowitz, *supra* note 1, at 63 n.290.

213. See, e.g., Kwestel & Seplowitz, *supra* note 1, at 63-65.

214. U.P.C. § 2-202(b)(2)(iii) (Supp. 1991).

knowledges that “such arrangements could, under the pre-1990 Code, have been used to deplete the estate and reduce the spouse’s elective-share entitlement.”²¹⁵ This is in stark contrast to the rationale for excluding insurance proceeds payable to a nonspouse under the prior Code. The former comment explained that insurance “is not ordinarily purchased as a way of depleting the probate estate and avoiding the elective share of the spouse.”²¹⁶ This about-face is telling not because it represents an inconsistency or contradiction, but because it indicates that as the types of nonprobate devices that may be used to disinherit a surviving spouse contract, ingenious decedents who wish to disinherit a spouse may begin to employ will substitutes that have not been used invidiously in the past. This danger is clearly visible in the redesigned UPC’s treatment of other testamentary substitutes.

c. *Testamentary substitutes includible if created during marriage*

With respect to other testamentary substitutes, the revised UPC does not fully promote the policies of the elective share. Section 2-202(b)(2)(iv),²¹⁷ with some modifications, follows the scheme of former section 2-202(l) which included in the augmented estate a number of

215. *Id.* § 2-202 comment.

216. U.P.C. § 2-202 comment (1983).

217. This subsection includes within the reclaimable estate:

[P]roperty transferred by the decedent to any person other than a bona fide purchaser at any time during the decedent’s marriage to the surviving spouse, to or for the benefit of any person, other than the decedent’s surviving spouse, if the transfer is of any of the following types:

- (A) any transfer to the extent that the decedent retained at the time of or during the two-year period next preceding his death the possession or enjoyment of, or right to income from, the property;
- (B) any transfer to the extent that, at the time of or during the two-year period next preceding the decedent’s death, the income or principal was subject to a power, exercisable by the decedent alone or in conjunction with any other person or exercisable by a nonadverse party, for the benefit of the decedent or the decedent’s estate;
- (C) any transfer of property, to the extent the decedent’s contribution to it, as a percentage of the whole, was made within two years before the decedent’s death, by which the property is held, at the time of or during the two-year period next preceding the decedent’s death, by the decedent and another, other than the decedent’s surviving spouse, with right of survivorship; or
- (D) any transfer made to a donee within two years before the decedent’s death to the extent that the aggregate transfers to any one donee in either of the years exceed \$10,000.

U.P.C. § 2-202(b)(2)(iv) (Supp. 1991).

transfers in the form of testamentary substitutes made during the decedent's marriage to the surviving spouse and for the decedent's benefit.²¹⁸

The first type of transfer is a retained life estate or the right to receive income or to possess or enjoy the transferred property within two years of death.²¹⁹ The next category includes any transfer in which the decedent retained a power in conjunction with another or in which a nonadverse party held a power over the income or principal for the decedent's benefit.²²⁰ This section includes a transfer in trust if the decedent retained a power to revoke or a power to invade the income or principal for his or her benefit. The third group includes a transfer in which the property is held within two years of the decedent's death by the decedent and a person other than the surviving spouse with a right of survivorship.²²¹ The last category consists of all gifts in excess of \$10,000 per donee per year made by the decedent within two years of death.²²² Like its predecessor, this section seeks to capture outright

218. This was true although the pre-1990 UPC, like the EPTL, included property owned outright by the decedent that was acquired prior to marriage in the augmented estate. *See N.Y. Est. POWERS & TRUSTS LAW § 5-1.1(c)* (McKinney 1981); U.P.C. § 1-201(11) (1983) (definition of "estate").

219. U.P.C. § 2-202(b)(2)(iv)(A) (Supp. 1991). The prior provision included only those transfers pursuant to which the decedent retained possession or enjoyment of or the right to income from the property at death. U.P.C. § 2-202(1)(i) (1983). This change closes a loophole that would have allowed a decedent to immunize a transfer by reserving the right to receive income until one month before death. *See I.R.C. § 2036(a)(1)* (1986); Kwestel & Seplowitz, *supra* note 1, at 58-59.

220. U.P.C. § 2-202(b)(2)(iv)(B) (Supp. 1991). Under the prior section, only those transfers were included in which the decedent retained a power over the principal at death. This section also did not explicitly include transfers if the power was held only by a nonadverse party. U.P.C. § 2-202(1)(ii) (1983). Thus, the revised version broadens and clarifies its predecessor. Nevertheless, whether this section includes transfers in which the decedent retained a power subject to an external standard (such as Horace's power in the trust he created in Case 1) is unclear. *See Kwestel & Seplowitz, supra* note 1, at 60-62 (arguing that transfers pursuant to which the decedent retained a power to invade should be includable in the augmented estate regardless of the existence of an external standard). Rejection of the external standard principle would accord with the UPC's policy of focusing on the decedent's benefit in the transferred property.

221. U.P.C. § 2-202(b)(2)(iv)(C) (Supp. 1991). The prior section applied to any transfer of property held at the decedent's death by the decedent and another with a right of survivorship. U.P.C. § 2-202(1)(iii) (1983). Thus, it was narrower than the current version because it did not contain the two year provision. It was also broader because it did not restrict the time of transfer to two years before the decedent's death. The more limited scope of the current provision can be explained by the addition of § 2-202(b)(2)(ii). *See supra* note 210.

222. U.P.C. § 2-202(b)(2)(iv)(D) (Supp. 1991). The prior section captured gifts in excess of \$3,000 per donee per year within two years of death. U.P.C. § 2-202(1)(iv) (1983). Excluding gifts below \$10,000 could potentially damage a surviving spouse in a less wealthy family. Some might argue that a better approach would be to include outright

gifts made in contemplation of death,²²³ but improves upon it by allowing a decedent to utilize the gift tax annual exclusion of section 2503(b) of the Internal Revenue Code without subjecting the gift to the right of election.²²⁴ Because the last category applies to outright gifts rather than testamentary substitutes with retained interests or powers, its restriction to transfers during marriage should not promote disinheritance. Although some applications of the doctrine of transfers in fraud of marital rights to antenuptial conveyances involve outright gifts,²²⁵ individuals are generally reluctant to dispose of property, even shortly before death, without retaining some power over or interest in the property.²²⁶

The comment to revised section 2-202 does not explain why testamentary substitutes, primarily life estates and transfers subject to retained powers, are includable only if created during marriage. It merely states that the purpose of including such transfers under the pre-1990 Code was "to protect the surviving spouse against so-called 'fraud on the spouse's share.'"²²⁷ Although it notes that the revisions "strengthen this feature of the pre-1990 Code," it refers only to general powers of appointment and life insurance.²²⁸ The comment's silence with respect to these testamentary substitutes suggests its endorsement of the prior rationale.

The justification given for this provision is to enable "a person to provide for children by a prior marriage, as by a revocable living trust,

transfers that exceed a given percentage, for example, five percent of the estate, but the author does not endorse this position. *See also* MINN. STAT. ANN. § 524.2-202(1)(iv) (West Supp. 1991) (the augmented estate includes any transfer by the decedent "made within one year of death of the decedent to the extent that the aggregate transfers to any one donee in the year exceeds \$30,000"); 20 PA. CONS. STAT. ANN. § 2203(a)(6) (Purdon Supp. 1991) (including property conveyed during marriage and within one year of death "to the extent that the aggregate amount so conveyed to each donee exceeds \$3,000").

223. This section appears to have been based upon I.R.C. § 2035 (1954), which included in the federal gross estate all of the decedent's gifts made within three years of death. The Economic Recovery Tax Act of 1981 substantially narrowed the scope of this section which, for the purposes of this Article, applies to the relinquishment within three years of death of powers retained pursuant to Internal Revenue Code §§ 2036, 2037, and 2038 and transfers of life insurance during that three year period. *See* I.R.C. § 2035 (1986).

224. The prior version captured gifts in excess of \$3,000 which negatively impacted on legitimate tax and estate planning. U.P.C. § 2-202(1)(iv) (1983). *See* Kwestel & Seplowitz, *supra* note 1, at 65-66.

225. In a significant number of cases involving what appeared to be an outright transfer, the decedent retained use of the property for life.

226. *See, e.g., In re Estate of Riefberg*, 58 N.Y.2d 134, 446 N.E.2d 424, 459 N.Y.S.2d 739 (1983) (amended shareholders agreement one day before death); *Newman v. Dore*, 275 N.Y. 371, 9 N.E.2d 966 (1937) (decedent transferred all of his real and personal property into a revocable trust three days before death).

227. U.P.C. § 2-202 comment (Supp. 1991).

228. *Id. See supra* notes 205, 212 and accompanying text.

without concern that such provisions will be upset by later marriage.”²²⁹ This rationale is untenable in light of the stated purpose of the prior section to capture in the augmented estate “transfers by the decedent during his lifetime which are essentially will substitutes, arrangements which give him continued benefits or controls over the property.”²³⁰ To the extent that a retained interest or power, such as the right to receive income from a trust or a power to revoke a trust, is equivalent to ownership of the property, this distinction does not fulfill the purpose of either the prior or current section.

Arguably, upsetting property arrangements made when the decedent did not contemplate marriage is unfair, particularly with respect to retained life estates. To give a surviving spouse a fractional share of the corpus and to diminish the share of the remaindermen (perhaps children from a prior marriage) may constitute too great a benefit to the spouse and too great a detriment to the children. After all, absent a power to invade the principal or a general presently exercisable power of appointment, a decedent with a reserved life estate has no lifetime interest in the principal, cannot undo the disposition unless all of the beneficiaries consent, and has no expectation at the time of the transfer that anyone other than the taxing authorities will be able to reach the corpus. Despite these objections, including reserved life estates in some manner in the elective share is the desirable approach and seems more consistent with the policy of marriage as a total partnership. By bringing the life estate into the marriage when a couple, such as Horace and Wanda, may enjoy its benefits, a spouse should be deemed to have converted it into partnership property. The right to enjoy possession of or the income from the property is so significant that the Internal Revenue Code approach equating this right with ownership should be followed with some modification.

In addition to the estate tax analogy, a creditor’s rights approach can be used to analyze the rights of a surviving spouse. Analogizing the survivor’s rights to those of a creditor of the decedent’s estate

229. U.P.C. § 2-202 comment (1983). The comment also indicates that “[t]he limitation to transfers during marriage reflects some of the policy underlying community property.” *Id.* This policy is also reflected in the following statutes: Mo. ANN. STAT. § 474.150(2) (Vernon Supp. 1990) (“Any conveyance of real estate made by a *married* person at any time without the joinder . . . of his spouse . . . is deemed to be in fraud of the marital rights of his spouse.”); Vt. STAT. ANN. tit. 14, § 473 (1989 & Supp. 1991) (right of election includes “voluntary conveyance by a husband of any of his real estate made during coverture and not to take effect until after his decease, and made with intent to defeat his widow in her claim to her share”); MODEL PROBATE CODE § 33(b) (1946) (“Any gift made by a *married* person within two years of the time of his death is deemed to be in fraud of the marital rights of his surviving spouse, unless shown to the contrary.”).

230. U.P.C. § 2-202 comment (1983).

produces a different result with respect to retained life estates. The estate tax analogy includes the property subject to the retained life estate as long as the decedent enjoyed it close to or until death. A creditor's rights analogy will require the creation of the life estate during marriage. Indeed, the UPC's scheme of including marital transfers of property subject to reserved life estates in the reclaimable estate can be justified by analogizing the surviving spouse to a creditor of the decedent's estate. At the time of such a transfer, a spouse knows that the surviving partner has a potential claim in the form of a right of election against the transferor's estate. Therefore, a spouse is on notice that any property transferred, particularly in the form of a testamentary substitute, may be subject to diminution to satisfy the elective share. This rationale should apply even though the surviving spouse is not technically the decedent's creditor during marriage. The surviving spouse, under the UPC or other elective share statutes, does not have a recognizable property interest in the decedent's assets before death, unlike common-law dower and community property systems.

The same analysis appears to be valid with respect to similar types of transfers in contemplation of marriage, yet it is not as compelling as if the prospective spouse had notice, but did not consent to the transfer. In that case, the individual who was unhappy with the disposition could decide not to marry. If the transfer occurred during the engagement period and the prospective spouse did not have notice, then this transfer can also be analogized to a transfer in fraud of creditor's rights. In both the engagement and marriage transfer cases, the decedent's motive for making the transfer may be questionable. Despite the rejection by the UPC and other progressive elective share statutes of the motive for the transfer in assessing the validity of the disposition, the underlying intent of these statutory schemes is to capture, in an objective sense, those devices that decedents can use for purposes of disinheritance.

Does this analysis mean that the UPC is correct in insulating from the right of election transfers of property subject to a retained life estate not made in contemplation of marriage, namely, the situation in Case 1? For the reasons already discussed, the UPC and similar elective share statutes should include such transfers because of the important benefits enjoyed by the decedent up to death. A partnership model focuses not on the time of transfer, but on the decedent's rights in the property existing during marriage and continuing immediately prior to death. Stated otherwise, property interests available to the partnership, as opposed to being generated during the partnership, should continue their characteristic as partnership property at death. Furthermore, the argument that a person, like Horace, who creates a reserved life estate prior to engagement, is not on notice of his future wife's (Wanda's) interest because he has not yet met her or contemplated marriage, is not per-

suasive. In this era of multiple marriages, the possibility of a future spouse should not be speculative.

Assuming that the partnership or marital-sharing theory is the correct or desirable model²³¹ does not mandate the exclusion of premarriage assets from the augmented estate. In a number of jurisdictions, all of a spouse's assets, regardless of when or how they were acquired, are subject to division in case of dissolution of the marriage by divorce.²³² Without incorporating the specific features of a particular equitable distribution system, this approach is equally reasonable in the elective share area and does not conflict with any of the justifications for the partnership theory of marriage.

This expansive view of partnership property does not mean that an elective share system such as the UPC must ignore the pre-existing property interests of the remaindermen, such as Doris and Sam in Case 1. Several alternatives exist to safeguard their interests in varying degrees. The approach that is potentially most detrimental to the remainder interests will embrace the federal estate tax model and give the surviving spouse the appropriate percentage of the principal under the UPC's accrual system or a fractional share under a more traditional elective share statute such as the EPTL or the Pennsylvania statute. A federal estate tax approach has the drawback of giving the surviving spouse an interest in the principal that the decedent lacked during life.

231. Some courts continue to view the elective share as principally serving the support needs of the surviving spouse. *See, e.g.*, *In re Merkel's Estate*, 190 Mont. 78, 82, 618 P.2d 872, 875 (1980) ("The primary purpose of the elective share statutes is to insure that the surviving spouse's needs are met, and that the spouse is not left penniless.").

232. *See, e.g.*, CONN. GEN. STAT. ANN. § 46b-81(c) (1986) (in McPhee v. McPhee, 186 Conn. 167, 440 A.2d 274 (1982), the court recognized the Connecticut statute subjecting premarital property to equitable distribution); IND. CODE § 31-1-11.5-11(b) (1988); KAN. STAT. ANN. §§ 23.201, 60.1610(b) (1988 & Supp. 1990) (upon dissolution, the court "shall divide the real and personal property of the parties, whether owned by either spouse prior to marriage, acquired by either spouse in the spouse's own right after marriage or acquired by spouse's joint efforts"); MASS. GEN. LAWS ANN. ch. 208, § 34 (West Supp. 1991); MICH. COMP. LAWS ANN. §§ 552.19, 552.23(1) (West 1988 & Supp. 1991); MONT. CODE ANN. § 40-4-202 (1989); N.H. REV. STAT. ANN. § 458: 16(a)(II)(m) (Supp. 1990); N.D. CENT. CODE § 14-05-24 (1981 & Supp. 1989) (in Fine v. Fine, 248 N.W.2d 838 (N.D. 1976), the equitable distribution statute allowed award of a spouse's separate property even if it was acquired before marriage, and the period when the property was acquired was only one factor to consider); S.D. CODIFIED LAWS ANN. § 25-4-44 (1984 & Supp. 1991); Vt. STAT. ANN. tit. 15, § 751 (1989); Wyo. STAT. § 20-2-114 (1977 & Supp. 1990). Despite the language in these statutes and the cases authorizing a division of property acquired before marriage or by gift or inheritance, commentators have noted a judicial reluctance to subject such property to equitable distribution "unless the circumstances warrant." *See J. OLDHAM, supra* note 194, at § 13.02[1][i]; Waggoner, *Spousal Probate Rights, supra* note 1, at 356 & n.41.

The second alternative would give the surviving spouse the same interest possessed by the decedent. Thus, in Case 1, Wanda would have the right to remain in the house for the rest of her life. This model is attractive, particularly with respect to the family residence, because it avoids displacing the survivor from the family home. Nevertheless, depending on the life expectancy of the surviving spouse, this alternative may seriously disadvantage the remaindermen's interests. Certainly, it would impinge on Horace's intent to provide for his children after his death. Under this alternative, Horace and his children might have been better off if he had not transferred the property, but had kept it in his probate estate.

The last alternative would give the surviving spouse a fractional share in the same property interest possessed by the decedent, an approach that follows section 2203(a)(2) of the Pennsylvania statute. Thus, in Case 1, Wanda would receive a fifty percent life interest in the house under the UPC's accrual system. This alternative has the advantage of benefitting the surviving spouse by recognizing the partnership nature of the life estate and giving the remaindermen an immediate interest in the property as well as leaving their interest in the corpus intact. The disadvantage would be to deprive the spouse of full enjoyment of the property which is a potentially serious detriment in the case of a family residence.

Of course, the easiest way for a decedent to avoid disadvantaging the remaindermen, who are frequently children from a prior marriage, is to disclose the existence of the arrangement to the prospective spouse who can then waive the right of election against such property or agree to a property settlement. This solution is also attractive because it encourages honesty between parties prior to marriage and avoids the troublesome issues associated with the doctrine of antenuptial transfers in fraud of marital rights regarding the sufficiency of notice to the prospective spouse.

If includability of antenuptial transfers in the augmented estate is to be judged by whether the surviving spouse had notice of the property arrangement prior to marriage, then unless the decedent informed the intended spouse of his or her possession of a life estate, the survivor should not be on inquiry notice. To the external world, the enjoyment of a life estate has many of the characteristics of full ownership. Consequently, the prospective spouse may not know whether the decedent's residence was owned as a life estate or in fee or whether an income stream was derived from a trust or an investment in the decedent's name. These reasons, coupled with the fact that in a substantial number of cases involving retained interests or powers, the decedent retained a life estate, suggest that an actual waiver requirement is desirable. Thus, if protection of the surviving spouse is a goal of the elective share, life

estates should be includable in the augmented estate regardless of when created unless the surviving spouse consented to the transfer.

Different considerations, however, are involved if the decedent retained a power to reach the transferred property. With respect to such transfers, the stated justification for former section 2-202 (which appears to have been adopted by the 1990 UPC) is also open to question.²³³ A person can provide for children of a prior marriage,²³⁴ insulated from any claim by a surviving spouse, by making an irrevocable gift or a transfer in trust without retaining any prohibited powers.²³⁵ If the individual has, as the comment to pre-1990 section 2-202 indicated, kept a power to invade the corpus, he or she has demonstrated an intent to retain substantial control over the property, including the power to "upset" the arrangement. Furthermore, because the *inter vivos* property arrangements and testamentary substitutes included in the reclaimable estate under section 2-202(b)(2)(iv) (or in the augmented estate under former section 2-202(1)) are primarily for the benefit of the decedent, the UPC's rationale for including only marital transfers is even more untenable. Therefore, such provisions should be subject to the right of election by a later spouse. To protect a trust with a retained power from the right of election in the event of a subsequent marriage, the settlor could relinquish the prohibited power prior to the marriage.²³⁶ The settlor would then be treated as making an irrevocable gift, and the transferred property would belong to the children free of the surviving spouse's claim under an elective share statute.²³⁷ Protecting children of

233. For views that the elective share should only reach property acquired during marriage, see OKLA. STAT. ANN. tit. 84, § 44(B) (West 1990) (elective share limited to property acquired by joint efforts after marriage); UTAH CODE ANN. §§ 75-2-201, -202 (1978 & Supp. 1991); Oldham, *supra* note 1, at 246 ("the forced share should be limited to the acquisitions of the decedent during marriage, other than gratuitous transfers from third parties").

234. The comment does not address the case of prenuptial testamentary substitutes for the benefit of people other than children or why such transfers should be excluded.

235. An example of this includes a power to revoke or invade the principal.

236. If the settlor dies within two years after the gift was made, the transferred property (even if in excess of \$10,000 "to any one donee in either of the years") is not subject to a right of election in a UPC jurisdiction. U.P.C. § 2-202(b)(2)(iv)(D) (Supp. 1991). Although the gift would be deemed to have been made in contemplation of death, it was not made during marriage. See 20 PA. CONS. STAT. ANN. § 2203(a)(6) (Purdon Supp. 1991) (gifts made during marriage and within one year of death are subject to the right of election). The same rationale applies to a release of a life estate or a prohibited power. See U.P.C. § 2-202(b)(2)(iv) (Supp. 1991). To the extent that gifts in contemplation of death are includable in the net estate for elective share purposes, property subject to prohibited powers which have been relinquished within one or two years of death should also be included. See I.R.C. §§ 2035(d)(2), 2038(a)(1) (1986).

237. See IND. CODE § 29-1-3-1 (1988) (adjusting the elective share where there are

a prior marriage can be accomplished by giving a surviving spouse the right to a smaller fractional interest in testamentary substitutes created prior to marriage. The decedent could also have entered into an agreement with the intended spouse that provides for a waiver of the right of election over such property.²³⁸

The retained powers section of UPC section 2-202 includes revocable trusts created by the decedent. Thus, the principal is includible in the augmented estate only if the trust was created during marriage. To be internally consistent, the UPC should add the principal of the trust to the augmented estate regardless of when the decedent established it. A trust subject to a power of revocation is similar to the property arrangements set forth in subsections (b)(2)(i), (ii), and (iii). The decedent who had the ability during life to obtain full ownership over property, has the same power under a revocable trust. A simple exercise of the power recaptures the principal. The same approach can be taken toward transfers of property when the decedent possessed a power to revoke or terminate because he or she was the only person with an interest in the property. Under the UPC's framework, if the decedent retained a power to revoke or terminate with an adverse party, the trust would then be a transfer of the type covered by subsection (b)(2)(iv) and would be includible only if created during marriage.

Ideally, trust property in which the decedent retained a power, either alone or in conjunction with another person, to revoke or consume the principal for his or her benefit should be includible in the augmented estate to the extent of the power. Regardless of whether the trust was created during the marriage, during the engagement, or before the engagement, the decedent had the ability to regain full ownership of all or part of the property. This same conclusion is reached by applying an estate tax analogy that examines the decedent's economic power over property at death.²³⁹ If the surviving spouse is viewed as a creditor, the spouse should be able to elect against the property if the decedent retained a power to revoke until death.²⁴⁰

d. Implications of the UPC's treatment of antenuptial transfers

A recent Arkansas case presents a stark illustration of the inequity that can be created under the UPC's scheme of excluding testamentary

children from a prior marriage); Wyo. STAT. ANN. tit. 2, § 2-5-101 (1980) (adjusting the elective share where there are children from a prior marriage). See also Wis. STAT. ANN. § 861.17(2)(West 1991) (an arrangement before or within one year of marriage to provide for issue of a prior marriage is not fraudulent).

238. The validity of these waivers is recognized in redesigned U.P.C. § 2-204.

239. See I.R.C. § 2038(a)(1) (1986).

240. See, e.g., N.Y. EST. POWERS & TRUSTS LAW § 10-10.6 (McKinney 1967) (treating settlor who reserved "an unqualified power of revocation" as "the absolute owner" of the transferred property for creditors' rights purposes).

substitutes created prior to marriage even if for the benefit of children from a previous marriage. In *Efird v. Efird*,²⁴¹ a man conveyed four lots to his children from a prior marriage and reserved a life estate. The transfer occurred two days before his second marriage without consideration or notice to his prospective wife.²⁴² Fourteen years after the transfer and one year after she learned of the conveyance, the wife sought the cancellation of the deeds as fraudulently denying her rights to the property. At trial, the wife testified that she and her husband were married and lived in the home that her husband built on the transferred property and that she decorated.²⁴³ When her husband became ill, she asked him "if she would have a place to live if something happened to him, and he said: 'No, you'll have to get on welfare, or your son'll have to take care of you.'"²⁴⁴ The wife was told that she had no interest in the house.

The appellate court held that the wife was entitled to a life estate in the property.²⁴⁵ It relied on a long line of Arkansas cases that applied the general rule:

[I]f a man or woman convey away his or her property for the purpose of depriving the intended husband or wife of the legal rights and benefits arising from such marriage, equity will avoid such conveyance or compel the person taking it to hold the property in trust for or subject to the rights of the defrauded husband or wife.²⁴⁶

241. 31 Ark. App. 190, 791 S.W.2d 713 (1990).

242. In fact, the prospective husband executed a quitclaim deed the following day (one day prior to the marriage) to correct the earlier warranty deed. *Id.* at 191, 791 S.W.2d at 713.

243. The husband died following the chancellor's dismissal of the wife's complaint; however, the case, which was also brought against the children, proceeded to appeal. *Id.*

244. *Id.* at 192, 791 S.W.2d at 714. Following this conversation, the wife checked the courthouse records and confirmed that she had no interest in the house. Evidently, the court assumed that by recording the deeds, the husband did not place his wife on constructive notice.

245. *Id.* at 195, 791 S.W.2d at 716. Because the wife testified at trial that she only wanted a life estate in the property, the trial court, on the defendant's motion, conformed the pleadings to her testimony.

246. *Id.* at 192, 791 S.W.2d at 714 (quoting *West v. West*, 120 Ark. 500, 504, 179 S.W. 1017, 1018 (1915)). Except for *West*, in which the transfer was upheld because the man apparently conveyed the property prior to his engagement, all other cited cases applying the rule held the transfer to be fraudulent. See *Wilhite v. Wilhite*, 242 Ark. 705, 415 S.W.2d 44 (1967); *Harrison v. Harrison*, 198 Ark. 64, 127 S.W.2d 270 (1939); *Roberts v. Roberts* 131 Ark. 90, 198 S.W. 697 (1917). The *Efird* court did not state what rights were defrauded, but presumably the court was referring to rights of dower, homestead, and a forced share. See also 20 ARK. STAT. ANN. §§ 28-11-301, 28-11-305, 28-11-307, 28-39-102, 28-39-401 (1987).

Although the husband appeared to be motivated by a desire to protect his children,²⁴⁷ the court stated that the wife need only establish constructive fraud which did not require an actual intent to deceive.²⁴⁸ Based on the evidence, the court had no difficulty finding constructive fraud.

The court in *Efird* did not justify its result based upon the nature of the marital relationship. The court was swayed by the short period of time between the transfer and the marriage as well as the fact that the wife lived on the property with her husband for thirteen years.²⁴⁹ The court also considered that the husband did not want her to have any interest in the property, but it did not discuss whether the house should be considered assets of the marital partnership.²⁵⁰ Such an analysis would, however, produce an affirmative response. Although the husband purchased the property with his own funds, his wife contributed to its furnishing and maintenance. In addition, the husband made mortgage payments during the marriage with funds that may have constituted marital assets.²⁵¹ Based on a marital-sharing or a return-of-contribution theory (both of which would recognize the wife's justifiable expectations), the property should be subject to the wife's right of election.²⁵² The same result would be reached under a commitment theory.

247. In each of the cases cited by the court, the husband transferred property to his children from a prior marriage or, as in *Roberts*, to a foster child, and reserved an actual or de facto life estate in himself. Except for *West v. West*, in which the court intimated that a provision for children might mitigate against finding a fraudulent intent, none of the cases recognized the exception that a transfer for the benefit of children is not fraudulent. *See also* *Barnett v. Barnett*, 209 Ark. 973, 193 S.W.2d 319 (1946) (invalidating a deed executed three days before a five week marriage which reserved a de facto life estate in favor of a nonmarital child insofar as dower and homestead rights were concerned). Apparently, only one reported Arkansas case involving the application of the doctrine to an antenuptial transfer made after an engagement failed to invalidate the conveyance. In *O'Connor v. Patton*, 171 Ark. 626, 286 S.W. 822 (1926), a man transferred real estate to a friend 11 days before marriage. The deed was absolute on its face, but the evidence indicated that the grantee held it on an oral trust for the husband's benefit. Although his motive was to prevent his future wife from acquiring any rights to the property, the court refused to set aside the deed because of equitable considerations. The wife was married at the time of her engagement and a few days after the wedding she refused to live with her new husband, indicating that she married solely for financial gain.

248. *Efird v. Efird*, 31 Ark. App. 190, 192, 791 S.W. 713, 715 (1990). *See Barnett*, 209 Ark. at 974, 193 S.W.2d at 319-20 (antenuptial deed is fraudulent even if the prospective husband lacked intent to disinherit his future wife). *But see West v. West*, 120 Ark. 500, 504, 179 S.W. 1017, 1018 (1915) ("Fraud is never presumed, but must be proved.").

249. *Efird*, 31 Ark. App. at 195, 791 S.W. at 715.

250. *Id.*

251. *See Wilhite v. Wilhite*, 242 Ark. 705, 707, 415 S.W.2d 44, 45 (1967) (a deed executed by a prospective husband 10 days before marriage held to be fraudulent because the wife "worked and helped to pay off a mortgage on the land after they were married").

252. These two notions underlie the partnership theory of marriage propounded by

Most importantly, the husband, in effect, transformed the property into marital property by making it the family home.²⁵³ In reality, he was no less the owner than if he had not transferred the property to his children because his lifetime enjoyment of it was basically comparable to full ownership. He reaped the same lifetime benefits from the property regardless of whether he transferred the remainder interest before or after his marriage or whether he made the transfer. Indeed, the husband's enjoyment of the property would not have been diminished if he executed the deed prior to his engagement or prior to meeting his second wife. In addition, his wife's contribution to the maintenance of the house in any of these cases would not have been substantially different. Based on a partnership theory of marriage, a determination of the includability of the property in the elective share should focus on the decedent's lifetime interest, not on the date a transfer may have occurred. Neither the doctrine of antenuptial transfers in fraud of marital rights nor the UPC produces an equitable or desirable result.

Despite the compelling reasons for giving the wife in *Efird* an interest in the house after her husband's death, under UPC section 2-202(b)(2)(iv)(A), she could not elect against it solely because the life estate was created before marriage. This result is untenable under either a partnership or support theory of the elective share. This transfer cannot be upheld as a conveyance designed to enable children from a prior marriage to enjoy the fruits of premarriage assets. The judgment in *Efird* protected the children from a prior marriage and the surviving spouse. Granted, the children will have to wait longer for their remainders to become possessory, but they have not lost their property interests. Furthermore, if the court awarded the widow an elective share, presumably she would have received only a fraction of the value of the property or the life estate. This result is preferable to the totally exclusionary system of the UPC and other elective share statutes such as the Pennsylvania and New York statutes. With the availability of the life estate (or any other testamentary substitutes) to the husband during marriage, the property begins to lose its characteristic of premarital or separate property, thereby diminishing the children's equities in it. As the length of the marriage increases and the commitment of the spouses

the UPC. See U.P.C. art. II, pt. 2 general comment (Supp. 1991). See also *supra* notes 184-87 and accompanying text.

253. This case demonstrates the continuing importance of the support justification for the elective share. See U.P.C. art. II, pt. 2 general comment (Supp. 1991). The importance of the family home to the surviving spouse and dependent children is also recognized in some jurisdictions (including Arkansas) by the homestead allowance. Presumably, if the husband had not transferred the property reserving a life estate, his surviving spouse could have claimed a homestead allowance.

to each other presumably increases, the children's claims decrease and the UPC's accrual system appropriately accommodates the competing interests of the surviving spouse and children from a prior marriage.

The equities of the children or other third parties are much greater when the decedent makes an outright transfer prior to marriage. For example, if Mr. Efird (or Horace, in Case 1) transferred the real property outright to his children, the UPC correctly excludes that property from the augmented estate. Property that was not brought into the partnership does not become partnership property. This situation, however, points to a serious shortcoming of the common-law doctrine applicable to outright as well as retained interest transfers. The common-law doctrine was only applicable to transfers in contemplation of marriage. Otherwise, the undesirable result of making outright transfers susceptible to challenge would ensue.²⁵⁴ Although outright transfers may on rare occasions pose a threat to the surviving spouse, they can best be dealt with under the applicable law of fraud.

V. RECOMMENDATIONS FOR CHANGE

Analysis of the common-law doctrine of antenuptial transfers in fraud of marital rights and several of the more progressive statutory elective share schemes indicates that integration of the two systems is necessary to promote the goals of marital partnership and support of the surviving spouse. To achieve more predictable results, consistency among the jurisdictions, and certainty in legitimate estate planning, it is desirable to abolish the common-law doctrine and amend the elective share statutes to take antenuptial transfers into account.²⁵⁵ Specifically, the following guidelines for statutory change are proposed.

254. *But see* Patecky v. Friend, 220 Or. 612, 350 P.2d 170 (1960) (granting a second wife a forced share despite a will contract requiring the decedent to leave his property to his first wife when the second wife did not have notice of the will contract or existing mutual wills).

255. The Advisory Committee Report reflects a step in this direction with respect to revocable testamentary substitutes. Macdonald's Suggested Model Decedent's Family Maintenance Act applies only to postnuptial transfers, "but the underlying policy applies also to antenuptial transfers." W. MACDONALD, *supra* note 33, at 359. This proposed legislation, which is concerned with providing sufficient maintenance for the surviving spouse, supports an inquiry into the size of the transfer prior to marriage. Macdonald believed that many cases show that courts applied the same illusory transfer doctrine used to test the validity of postnuptial transfers to prenuptial transfers. *Id.* at 360 n.25. This is particularly true with personality because the right to dower supported actions to set aside conveyances of realty. To determine whether a prenuptial transfer should be subject to the surviving spouse's right of election, he proposed using a "reasonableness" test. *Id.* at 363-64. "Recovery would not hinge on the state of mind of either the man or the woman at the time of transfer, or whether the husband is alive or dead at the time of the litigation, on the fortuitous nature of the wife's 'interest' during coverture, or on the type of property that is involved." *Id.* at 364-65.

A. *Outright Transfers Before Marriage*

Under the common-law doctrine or its codifications, outright transfers can be divided into two types: those that occur during the engagement period or in contemplation of marriage and those that precede marriage plans. As a general rule, the doctrine only applies to the first category. Under the statutory approaches utilizing objective criteria to determine the includability of *inter vivos* dispositions in the elective share, outright transfers of either type are excluded. In Pennsylvania or a UPC jurisdiction, only outright transfers in contemplation of death and during marriage constitute testamentary substitutes. New York excludes outright transfers from the net estate, but even revocable gifts *causa mortis* may be excluded if given before marriage.²⁵⁶

In contemporary society, couples generally do not enter into property settlements upon engagement. Abolishing the distinction between transfers before and after engagement is recommended because the actual date of the engagement may be uncertain.²⁵⁷ (This also eliminates the need to determine if a person transferred property in contemplation of an engagement.) Dispensing with this distinction produces an analytically correct result because it no longer requires a finding of marital rights prior to marriage when none exist. Particularly in jurisdictions that no longer recognize common-law dower or that have not adopted a community property system, it is clear that inheritance is not a vested or contingent property right. A spouse may make large outright gifts of property during marriage²⁵⁸ free from any claim of the surviving spouse.²⁵⁹ The only exceptions are transfers or gifts made in contemplation

256. See *supra* note 170 (changes recommended by the Advisory Committee Report).

257. Uncertainty as to whether the transfer occurred before or after the engagement was a factor in the court's decision to order a remand in *Papouchis v. Papouchis*, 28 A.D.2d 554, 280 N.Y.S.2d 160 (1967). See WIS. STAT. ANN. § 861.17 (West 1991) (applying to transfers in contemplation of marriage).

258. An argument can be made that all post-marriage transfers to nonfamily members without adequate consideration, regardless of whether they are outright, should be included in the augmented estate. This approach is a modification of the community property model.

259. The only exception is for outright gifts made in contemplation of death. Under the UPC, the surviving spouse has a right to elect against any transfers made by the decedent within two years of death "to the extent that the aggregate transfers to any one donee in either of the years exceeded \$10,000." U.P.C. § 2-202(b)(2)(iv)(D) (Supp. 1991). Pennsylvania's analogous provision applies only to transfers made within one year of death. 20 PA. CONS. STAT. ANN. § 2203(a)(6) (Purdon Supp. 1991). In New York, a surviving spouse may elect against a lifetime gift only if it is a gift *causa mortis*. N.Y. EST. POWERS & TRUSTS LAW § 5-1.1(b)(1)(A) (McKinney 1981). But see *supra* note 170 (recommendation of the Advisory Committee Report).

of death under a statute like the UPC or the Pennsylvania estates law. Finally, elimination of the doctrine of antenuptial transfers in fraud of marital rights should not decrease the level of protection afforded to the surviving spouse or increase the number of cases of spousal disinheritance. It is unlikely that a decedent will give away all or a significant part of his property prior to death,²⁶⁰ albeit prior to marriage.

The vast majority of cases dealing with premarriage transfers involve a retained interest or power, generally a reserved life estate. Therefore, without significantly harming a surviving spouse, an elective share statute should exclude outright transfers from the right of election unless they are made in contemplation of death. To avoid upsetting outright gifts, an elective share statute should consider shortening the period within which such gifts are captured to one year. Any transfer in excess of \$10,000 to any donee within one year of death would be includible in the augmented estate, regardless of whether it preceded or followed marriage. Such measures cover transfers motivated by a desire to disinherit a surviving spouse or that unfairly benefit children from a prior marriage. Under this approach, difficult tracing problems are avoided, and legitimate estate and tax plans remain intact.

B. Will Substitutes

For the foregoing reasons, the distinction between will substitutes made during the engagement period and those not made in contemplation of marriage should be abolished. Any will substitute created by a decedent that confers a substantial benefit or gives the grantor significant economic control should be subject to the right of election whether made before or during marriage. The UPC's reclaimable estate should be amended to include not only property over which a decedent holds a presently exercisable power of appointment, unilaterally severable interests held by a decedent with right of survivorship, and life insurance regardless of when these interests arose, but also the other categories of will substitutes enumerated in section 2-202(b)(2)(iv). The surviving spouse should have, at a minimum, a fractional share of retained life estates created prior to marriage either by the enjoyment of the property or an income stream for life. A similar approach should be taken with respect to the Pennsylvania and New York elective share statutes. This yields a consistent result in Cases 1, 2, and 3 and equitably accommodates the interests of the surviving spouse and any children from a prior marriage.

260. See *supra* note 226 and accompanying text.

VI. CONCLUSION

The proposed guidelines, if adopted, advance fairness, predictability, and consistency of elective share statutes by promoting the goal of marriage as a total partnership. The doctrine of antenuptial transfers prior to marriage is rejected. However, the guidelines incorporate the underlying theory of the doctrine by recognizing and counteracting the detrimental effect many premarriage transfers have on the rights of the surviving spouse.

Police Power Absolutism and Nullifying the Free Exercise Clause: A Critique of *Oregon v. Smith*

JOHN DELANEY*

“A regulation neutral on its face may, in its application, nonetheless offend . . . the free exercise of religion.”

Chief Justice Warren Burger¹

INTRODUCTION

In the recent landmark decision of *Oregon v. Smith*,² the United States Supreme Court, in Justice O’Connor’s words, “dramatically departs from well-settled First Amendment jurisprudence”³ in a “paradigm free exercise”⁴ case by reframing a core dimension of free exercise doctrine. Justice Blackmun calls this reframing “a wholesale overturning of settled law concerning the Religion Clauses of our Constitution.”⁵ A bare Court majority severely restricted the scope of the application of the traditional compelling interest balancing test for assessing claims of a burdening of religious practice and provided no real substitute test. What it sacrificed is elegantly captured by Justice O’Connor’s description of the central significance of this balancing test in implementing the free exercise clause:

The compelling interest test effectuates the First Amendment’s command that religious liberty is an independent liberty, that it occupies a preferred position, and that the Court will not permit encroachments upon this liberty, whether direct or indirect, unless

* Associate Professor of Law, City University of New York Law School at Queens College; B.S.S., 1954, Fordham University; J.D., 1959, New York Law School; LL.M., 1965, New York University School of Law; M.A., 1972, New School for Social Research. I wish to thank the following persons for their assistance with this Article: my colleagues at the City University Law School, Susan Carpenter, Steve Loffredo, and Peter Margulies; Ali Khan of the Washburn University Law School; Judge Luke Charde; Jesse Kasowitz, Esq.; my students, Shawn Boatright, Gilma Camargo, David Campbell, Marsha Edwards, and Robert Perry; my secretary, Betty Tabor; and my wife Lisa Blitman, Esq. I am also indebted to the CUNY Law School, its deans, and faculty for supporting this research and writing with a semester of paid leave.

1. *Wisconsin v. Yoder*, 406 U.S. 205, 220 (1972).
2. 110 S. Ct. 1595 (1990).
3. *Id.* at 1606 (O’Connor, J., concurring).
4. *Id.* at 1612 (O’Connor, J., concurring).
5. *Id.* at 1616 (Blackmun, J., dissenting).

required by clear and compelling governmental interests "of the highest order." . . . The compelling interest test reflects the First Amendment's mandate of preserving religious liberty to the fullest extent possible in a pluralistic society.⁶

The immediate design of the Court majority is manifest — to foreclose free exercise analysis of any kind when a claim arises from the application of a typical facially neutral, generally applicable criminal statute. Yet, this drastic curtailing of free exercise claims may be part of a more resonant orchestration designed to refashion the role of the federal courts by leaving accommodation of claims that conflict with such statutes to the legislature with the likely result that recognition of minority religious claims will be held hostage to majoritarian politics.

Implicit in this startling dispatch to the legislature of free exercise claims is an alteration of our theory of democratic government which authorizes majoritarian decisionmaking within a framework of fundamental rights guaranteed by the Bill of Rights and protected by the courts. The Court is altering the proper relationship between individuals and government by enlarging state power through the curtailing of the occasions when individual constitutional rights may be considered by the courts. This is a "vision . . . bordering on the authoritarian . . . and insufficiently sensitive to human rights and needs."⁷ It is not surprising, therefore, that the *Smith* Court's curtailing of free exercise claims, including its restricted role for the federal courts and its refashioned democratic theory, has provoked vehement criticism from within Congress and from a diverse coalition of religious groups throughout the country. In addition, it has provoked a congressional bill to "protect the free exercise of religion" through the restoration of the compelling interest test.⁸

6. *Id.* at 1609, 1613 (O'Connor, J., concurring).

7. See L. TRIBE, AMERICAN CONSTITUTIONAL LAW ix (2d ed. 1988). Confirmation of this trend is proliferating. In *McCleskey v. Zant*, 111 S. Ct. 1454 (1991), the Court adopted new rules curtailing the number of times defendants sentenced to death can petition federal courts to consider new evidence or new arguments. In *Arizona v. Fulminante*, 111 S. Ct. 1246 (1991), the Court decided that a confession beaten out of a defendant may be only "harmless error" not warranting reversal of a conviction. In *Rust v. Sullivan*, 111 S. Ct. 1759 (1991), the Court, for the first time, approved executive agency censorship of what those who accept federal funds may say in carrying out their professional duties, in contrast to the traditional regulation of what they may do with those funds.

8. Numerous members of Congress have co-sponsored the Religious Freedom Restoration Act of 1990. H.R. 5377, 101st Cong., 2nd Sess. (1990). The authority for this bill is § 5 of the fourteenth amendment to the United States Constitution, which states that "[t]he Congress shall have power to enforce, by appropriate legislation, the

The claimants in *Smith* were denied unemployment insurance benefits after they were fired from their jobs as drug counselors for ingesting peyote in an annual sacramental rite of the Native American Church. They argued that the Court's rule and principle embedded in four prior unemployment cases controlled their unemployment case. More specifically, they argued that their dismissal fell within the free exercise exemption carved out by unemployment case precedents. Although the Oregon courts agreed, the *Smith* majority rejected this claim and reasoned that these cases did not apply to the claim because the denial of benefits resulted from behavior that fell within "an across-the-board criminal prohibition on a particular form of conduct."⁹ In addition, the Court emphatically rejected the application of the *Sherbert* compelling interest balancing test¹⁰ to the *Smith* facts and to *any* free exercise claim that

provisions of this article." U.S. CONST. amend. XIV, § 5.

The broad-based Coalition for the Free Exercise of Religion includes the American Jewish Congress, American Jewish Committee, Evangelical Lutheran Church in America, National Council of Churches, Baptist Joint Committee on Public Affairs, National Association of Evangelicals, Native American Church of North America, Union of American Hebrew Congregations, General Conference of Seventh Day Adventists, Agadeth Israel of America, American Civil Liberties Union, Church of the Brethren, and many other religious and secular groups. (Coalition for the Free Exercise of Religion, 200 Maryland Avenue, N.E., Washington, D.C.). Michael McConnell, a free exercise commentator, has expressed some of the concerns:

[A] requirement that all witnesses must testify to facts within their knowledge bearing on a criminal prosecution . . . if applied without exception, could abrogate the confidentiality of the confessional. Similarly, a general prohibition of alcohol consumption could make the Christian sacrament of communion illegal, uniform regulation of meat preparation could put kosher slaughterhouses out of business, and prohibitions of discrimination on the basis of sex or marital status could end the male celibate priesthood.

McConnell, *The Origins and Historical Understanding of Free Exercise of Religion*, 103 HARV. L. REV. 1410, 1419 (1990). The application of *Smith* has included the dismissal, "with deep regret," by a federal district court in Rhode Island of its prior holding that an autopsy should not have been performed on a son of Hmong parents over their religious objection. *Yang v. Sturner*, 750 F. Supp. 558, 558 (D.R.I. 1990). See also *Salvation Army v. Department of Community Affairs*, 919 F.2d 183, 196 (3d Cir. 1990) (state civil statute and regulations governing boarding homes "are not specifically addressed to religious practice and therefore, under . . . *Smith* are not susceptible to a free exercise clause challenge"); *Montgomery v. County of Clinton*, 743 F. Supp. 1253 (W.D. Mich. 1990) (autopsy performed on man over religious objection of Jewish mother did not violate her free exercise rights). An Ohio appellate court described the free exercise clause after *Smith* as "a puff of smoke." *State v. Flesher*, No. 89-P-2084, 1990 WL 73953, at 4 (Ohio Ct. App. June 1, 1990). For Jesse H. Choper, Dean and Professor of Law at the University of California at Berkely, *Smith* represents "the demise of the Free Exercise Clause." See 59 U.S.L.W. 2272, 2274 (Nov. 6, 1990) (synopsis of the Constitutional Law Conference held September 14-15, 1990).

9. Employment Div. v. Smith, 110 S. Ct. 1595, 1603 (1990).

10. *Sherbert v. Verner*, 374 U.S. 398 (1963).

arises from the application of “an across-the-board criminal prohibition.”¹¹ The chief rationale for repudiating the validity of applying the *Sherbert* balancing test to the *Smith* claim was explicitly stated by the Court:

The *Sherbert* test . . . was developed in a context that lent itself to individualized governmental assessment of the reasons for the relevant conduct. . . . [A] distinctive feature of unemployment compensation programs is that their eligibility criteria invite consideration of the particular circumstances behind an applicant’s unemployment. . . . “The ‘good cause’ standard [in unemployment compensation statutes] created a mechanism for individualized exemptions. . . .” [O]ur decisions in the unemployment cases . . . have nothing to do with an across-the-board criminal prohibition on a particular form of conduct.¹²

The first thesis of this Article is that this rationale is mistaken and has led the Court to distinguish incorrectly between types of cases in which the Court must weigh the competing interests of the individual and the government. It is simply untrue that the rationale for balancing these interests in unemployment cases does not apply when a criminal statute prohibits the conduct. The historical and everyday reality is that both realms of law manifestly require “individualized governmental assessment of the reasons for the relevant conduct.”¹³ The core criminal law principles of *mens rea* and *actus reus* mandate, not simply invite, “consideration of the particular circumstances behind” a criminal defendant’s external behavior.¹⁴ These animating principles are detailed in an array of criminal law defenses which for centuries have indisputably comprised “a mechanism for individualized exemptions.”¹⁵

The Court’s error in *Smith* is fundamental and undermines the major premise that drives the majority’s analysis: the attribution of “talismanic”¹⁶

11. *Smith*, 110 S. Ct. at 1603.

12. *Id.* (quoting *Bowen v. Roy*, 476 U.S. 693, 708 (1986)) (opinion of Burger, C.J., joined by Powell and Rehnquist, J.J.). The requirement of state authorized individualized exemptions is also embodied in *Bowen*: “If a state creates such a mechanism, its refusal to extend an exemption to an instance of religious hardship suggests a discriminatory intent.” *Bowen*, 476 U.S. at 708. See also *Hobbie v. Unemployment Appeals Comm’n*, 480 U.S. 136, 142 n.7 (1987) (Florida’s statutory unemployment scheme, “which labels and penalizes behavior dictated by religious beliefs as intentional misconduct, exhibits greater hostility toward religion than one deeming such resignations to be ‘without good cause.’”).

13. *Smith*, 110 S. Ct. at 1603.

14. *Id.*

15. *Bowen v. Roy*, 476 U.S. 693, 708 (1986).

16. In disagreeing with the Court’s failure to apply the compelling interest test,

power to an ordinary across-the-board criminal statute, a power that bars a claim that the free exercise of religion has been burdened when the claim arises from the application of a criminal statute. The Court's majority relied on this flawed premise in rejecting "a settled and inviolate principle of [the] Court's First Amendment jurisprudence" in the criminal context.¹⁷ This principle is that the test of the "constitutionality of a state statute that burdens the free exercise of religion" requires that both the "law in general, and the State's refusal to allow a religious exemption in particular, are justified by a compelling interest that cannot be served by less restrictive means."¹⁸

The second thesis of this Article is that the Court's major premise in *Smith* is wrong on two counts. The Court's premise violates core principles expressed in our theory of just punishment within a framework of constitutional criminal law and is also unjustified under our theory of crime control. The third and culminating thesis of this Article is that the Court's mistaken rationale and major premise result in a holding and principle that nullifies *de facto* the free exercise clause when a claim arises in the context of criminal law. The *Smith* Court's rejection of the command of the free exercise text is in favor of police power absolutism and is therefore antithetical to settled constitutional principles embedded in our jurisprudence.

After a detailed statement of the *Smith* facts and its complicated procedural history, Part I of this Article examines the Court's key rationale and major premise and explains that in order to understand this rationale and premise it is necessary to unfold the nature and scope of the governmental interest in across-the-board criminal statutes. In Part II, the Court's rationale and major premise are analyzed and critiqued from the standpoint of our theory of just punishment. In Part III, this rationale and premise are analyzed and critiqued from the perspective of our framework of constitutional criminal law. Lastly, in Part IV, the Court's rationale and premise are analyzed and critiqued from the standpoint of our theory of crime control.

This Article, however, does not systematically scrutinize other central contentions from *Smith* that are aptly analyzed and persuasively refuted in Justice O'Connor's concurring opinion, including the Court's "dis-

Justice O'Connor noted that there "is nothing talismanic about neutral laws of general applicability or general criminal prohibitions." *Employment Div. v. Smith*, 110 S. Ct. 1595, 1612 (O'Connor, J., concurring).

17. *Id.* at 1616 (Blackmun, J., dissenting).

18. *Id.* at 1615 (footnote omitted). In addition to compelling interest, the Court has utilized other language without specifying any difference in doctrinal meaning. See, e.g., *United States v. Lee*, 455 U.S. 252, 257-58 (1982) ("overriding governmental interest"); *Wisconsin v. Yoder*, 406 U.S. 205, 215 (1972) ("interests of the highest order").

torted view”¹⁹ of free exercise precedents in favor of “the single categorical rule that ‘if prohibiting the exercise of religion . . . is . . . merely the incidental effect of a generally applicable and otherwise valid provision, the First Amendment has not been offended.’”²⁰ The Article also does not analyze the discussion in Justice Blackmun’s eloquent dissent of the remarkably special facts regarding the religious interest presented by Smith’s and Black’s sacramental use of peyote in light of the precise nature of the state crime control interest at stake.²¹ The *Smith* decision can, and should, be critiqued from a variety of perspectives. My primary interest is not in the Court’s resolution of the facts, but rather in its sweeping principle, the drastic curtailing of the scope of the application of the compelling interest balancing test as a means to blunt the application of the free exercise clause and the impact of this principle at a time when America is becoming more religiously diverse as immigrants flow into the United States from Asia, Africa, the Middle East, and Latin America.²²

The reframing of free exercise jurisprudence by the Supreme Court arises from the firing of respondents Alfred Smith and Galen Black from their jobs as counselors for a private drug rehabilitation organization in Oregon. Smith and Black were members of the Native American Church who ingested peyote as part of an annual sacramental rite. When they applied for unemployment compensation benefits, the state agency denied their applications on the basis that they were discharged for work-related “misconduct,” a disqualifying test specified in the Oregon unemployment compensation statute. The Oregon Court of Appeals

19. *Smith*, 110 S. Ct. at 1616 (Blackmun, J., dissenting).

20. *Id.* at 1607 (O’Connor, J., concurring).

21. See *id.* at 1616-21.

22. [M]ore than two million immigrants [have arrived] across the Pacific Ocean in the past decade. . . . In 1965, Asian-Americans numbered barely one million; a quarter of a century later, the census counted 7.3 million Americans of Asian and Pacific Islander background. The Korean-American population has risen to nearly 800,000 and the East Indian population stands at 815,000. The Vietnamese population in the United States is about 615,000, Laotians number 149,000, Cambodians 147,411, Thais 91,000, Hmong 90,000 and Pakistanis 81,000.

N.Y. Times, June 12, 1991, at A1, col. 1. See also Pipes, *The Muslims Are Coming! The Muslims Are Coming!*, NAT’L REV., Nov. 10, 1990, at 28, 31 (“Muslims total two to three million in the United States. . . . Muslims will become the second largest religious community in about ten years.”). Michael McConnell depicts the importance of the free exercise clause for “unfamiliar faiths”: “One rarely sees laws that force mainstream Protestants to violate their consciences. Judically enforceable exemptions under the free exercise clause are therefore needed to ensure that unpopular or unfamiliar faiths will receive the same consideration afforded mainstream or generally respected religions by the representative branches.” McConnell, *supra* note 8, at 1419-20.

reversed the agency's determination and held that the denial of benefits violated their free exercise rights as guaranteed by the first amendment.²³

On appeal to the Oregon Supreme Court, the agency argued that the denial of benefits was permissible because consumption of peyote is a crime under Oregon law. The Oregon Supreme Court rejected the relevance of this argument, affirmed the judgment of the Oregon Court of Appeals, and held that "the legality of ingesting peyote does not affect our analysis of the state's interest. The state's interest . . . must be found in the unemployment compensation statutes, not in the criminal statutes proscribing the use of peyote."²⁴ When the case was first reviewed by the United States Supreme Court in 1988 (*Smith I*), the Court determined that the alleged illegality of the respondents' peyote consumption was relevant to the constitutional analysis.²⁵ The Court remanded the case to the Oregon Supreme Court to resolve the question of the legality of the religious use of peyote in Oregon.

On remand, the Oregon Supreme Court, in a *per curiam* opinion, reaffirmed its prior decision as to the intent of the Oregon legislature: "[I]t was immaterial to Oregon's unemployment compensation law whether the use of peyote violated some other [criminal] law . . . [because] [t]he state's interest is simply the financial interest in the payment of benefits from the unemployment insurance fund to this claimant."²⁶ The Oregon Supreme Court also resolved the question posed on remand by the United States Supreme Court and stated that, "the Oregon statute against possession of controlled substances, which include peyote, makes no exception for the sacramental use of peyote."²⁷ The Oregon Supreme Court reaffirmed its prior holding that Smith and Black were entitled to unemployment compensation benefits because "outright prohibition of good faith religious use of peyote by adult members of the Native American Church would violate the First Amendment directly and as interpreted by Congress."²⁸

The Supreme Court emphatically rejected this proposition as applied to Smith and Black in the second appearance of *Smith* in the United States Supreme Court.²⁹ The Court articulated a powerful new holding

23. *Black v. Employment Div.*, 75 Or. App. 735, 707 P.2d 1274 (1985) (en banc), *aff'd*, 301 Or. 221, 721 P.2d 481 (1986), *vacated*, 485 U.S. 660 (1988).

24. *Smith v. Employment Div.*, 301 Or. 209, 219, 721 P.2d 445, 450 (1986), *vacated*, 485 U.S. 660 (1988).

25. *Employment Div. v. Smith*, 485 U.S. 660, 662 (1988).

26. *Smith v. Employment Div.*, 307 Or. 68, 71, 763 P.2d 146, 147 (1988), *rev'd*, 110 S. Ct. 2605 (1990).

27. *Id.* at 72-73, 763 P.2d at 148 (footnote omitted).

28. *Id.* at 73, 763 P.2d at 148.

29. *Employment Div. v. Smith*, 110 S. Ct. 1595, 1595 (1990).

and principle for the future. If the claim of a burdening of religious practice arises from the application of a facially neutral, generally applicable criminal statute, the burdening has an incidental effect and a first amendment violation does not exist.³⁰

I. THE *SMITH* PRINCIPLE UNFOLDED

A. *The Court's Rationale and Major Premise Detailed*

The key reason why the *Smith* Court rejected the application of the *Sherbert* balancing test is that the prior decisions of the Court that have embodied this balancing test "have nothing to do with an across-the-board criminal prohibition on a particular form of conduct."³¹ The

30. *Id.* at 1602-03. In this Article, the description "typical criminal statute" means an otherwise valid, facially neutral, generally applicable criminal statute.

31. *Id.* at 1603 (referring to *Hobbie v. Unemployment Appeals Comm'n*, 480 U.S. 136 (1987); *Thomas v. Review Bd.*, 450 U.S. 707 (1981); *Sherbert v. Verner* 374 U.S. 398 (1963)). The Court distinguished its other decisions in which the free exercise clause was applied in the context of facially neutral "across-the-board criminal prohibitions":

The only decisions in which we have held that the First Amendment bars application of a neutral, generally applicable law to religiously motivated action have involved not the Free Exercise Clause alone, but the Free Exercise Clause in conjunction with other constitutional protections, such as freedom of speech and of the press, see *Cantwell v. Connecticut*, 310 U.S., at 304-307, 60 S. Ct., at 903-905 (invalidating a licensing system for religious and charitable solicitations under which the administrator had discretion to deny a license to any cause he deemed nonreligious); *Murdock v. Pennsylvania*, 319 U.S. 105, 63 S. Ct. 870, 87 L. Ed. 1292 (1943) (invalidating a flat tax on solicitation as applied to the dissemination of religious ideas); *Follett v. McCormick*, 321 U.S. 573, 64 S. Ct. 717, 88 L. Ed. 938 (1944) (same), or the right of parents, acknowledged in *Pierce v. Society of Sisters*, 268 U.S. 510, 45 S. Ct. 571, 69 L. Ed. 1070 (1925), to direct the education of their children, see *Wisconsin v. Yoder*, 406 U.S. 205, 92 S. Ct. 1526, 32 L. Ed. 2d 15 (1972) (invalidating compulsory school-attendance laws as applied to Amish parents who refused on religious grounds to send their children to school). Some of our cases prohibiting compelled expression, decided exclusively upon free speech grounds, have also involved freedom of religion.

Id. at 1601 (footnote omitted). The Court majority concluded that the *Smith* facts are distinguishable because "[t]he present case does not present such a hybrid situation, but a free exercise claim unconnected with any communicative activity or parental right." *Id.* at 1602.

Justice O'Connor, in her concurring opinion, replied that "[t]he Court endeavors to escape from our decisions in *Cantwell* and *Yoder* by labeling them 'hybrid' decisions . . . but there is no denying that both cases expressly relied on the Free Exercise Clause . . . and that we have consistently regarded those cases as part of the mainstream of our free exercise jurisprudence." *Id.* at 1609 (O'Connor, J., concurring).

Let us assume, arguendo, that the *Smith* Court's description of these decisions is

Court stated that “the conduct at issue in those cases was not prohibited by law.”³² Justice Scalia distinguished these “balancing test” decisions from the *Smith* facts and dismissed the claim of a burdening of religious practice on the grounds that this test “was developed in a context that lent itself to individualized governmental assessment of the reasons for the relevant conduct. . . . [The] eligibility criteria [of these statutes] invite consideration of the particular circumstances behind an applicant’s unemployment.”³³ The key unemployment compensation decisions, *Sherbert v. Verner*,³⁴ *Thomas v. Review Board*,³⁵ and *Hobbie v. Unemployment Appeals Commission*,³⁶ focused on the “good cause” standard for quitting a job, which “created a mechanism for individualized exemptions.”³⁷

Thus, in Scalia’s words, “our decisions in the unemployment cases stand for the proposition that where the State has in place a system of individual exemptions, it may not refuse to extend that system to cases of ‘religious hardship’ without compelling reason.”³⁸ The Court, however, found that the balancing test and the proposition affirmed by these decisions “have nothing to do” with a claim of a burdening of religious practice arising from the application of an “across-the-board criminal prohibition.”³⁹ The Court did not explicitly acknowledge the manifestly

correct (which it is not). How does the descriptive reality of these cases translate into a normative principle that a free exercise claim is not cognizable if it is unconnected with any communicative activity or parental right? How is fidelity to one’s oath to support the Constitution satisfied by holding the free exercise text hostage to the existence of an additional constitutional text? More specifically, how is such fidelity satisfied by judicially adding to the first amendment a requirement that the free exercise claim must also be accompanied by a case-based principle which is derived from a different constitutional text (e.g., the *Pierce*-based right of parents to direct the education of their children derived from the liberty interest protected by the fourteenth amendment)? How is all of this squared with a judicial conservative’s commitment to the foundational text of the Constitution as the centerpiece of constitutional interpretation?

32. *Id.* at 1598.

33. *Id.* at 1603.

34. 374 U.S. 398 (1963).

35. 450 U.S. 707 (1981).

36. 480 U.S. 136 (1987). The *Smith* majority was mistaken in its statement that “[a]pplying [the *Sherbert*] test we have, on *three* occasions, invalidated state unemployment compensation rules that conditioned the availability of benefits upon an applicant’s willingness to work under conditions forbidden by his religion.” Employment Div. v. Smith, 110 S. Ct. 1595, 1602 (1990) (emphasis added). Actually, on four occasions, not three, the Court applied the *Sherbert* test in the unemployment insurance area. The fourth decision, by a unanimous Court, is *Frazee v. Illinois Department of Employment Security*, 489 U.S. 829 (1989). The Court overlooked the respondents’ reliance on *Frazee* in stating that “[r]espondents’ claim for relief rests on our decisions in [*Sherbert*, *Thomas*, and *Hobbie*.]” *Smith*, 110 S. Ct. at 1598.

37. *Bowen v. Roy*, 476 U.S. 693, 708 (1986).

38. *Smith*, 110 S. Ct. at 1603 (citing *Bowen v. Roy*, 476 U.S. 693, 708 (1986)).

39. *Id.*

contrary implications of this conclusion: that across-the-board criminal statutes have no comparable standard and no comparable mechanism for individualized exemptions. Thus, because a “system of individual exemptions”⁴⁰ is not in place, the *Sherbert* compelling interest test is inapplicable to the *Smith* claim of religious hardship. The test is inapplicable because the reason for the test does not apply.

The Court’s holding and principle flow from this rationale. If “a generally applicable and otherwise valid”⁴¹ across-the-board criminal prohibition exists, any impact on a person’s religious practice is “merely the incidental effect” of the application of the statute and “the first amendment has not been offended.”⁴² To trigger free exercise balancing analysis, the object of a criminal statute on its face must proscribe the exercise of religion.⁴³ Otherwise, any burdening of religious practice is automatically characterized as “merely the incidental effect” of the application of the statute, free exercise balancing is foreclosed and indeed, free exercise analysis is barred.⁴⁴

Hence, in a criminal prosecution, a prosecutor may refute a claim that a criminal charge burdens the free exercise of religion by simply responding that the object of the criminal statute is not to burden religious exercise and thus, any burden actually imposed is “merely the incidental effect” of an otherwise valid provision. Applying *Smith*, the singular role of the trial court will be to determine that the object of the statute on its face is not to burden religious exercise.

Such a finding is virtually certain, even pro forma, because as Justice O’Connor noted, it would be an “extreme and hypothetical situation in which a State directly targets a religious practice”⁴⁵ on the face of a criminal statute. The reason underlying this conclusion is that such direct targeting by a legislature of religious practice violates the respect for the free exercise clause that is embedded in our political and civil culture. Once the trial court makes its virtually certain finding that the criminal statute does not facially burden religious exercise, any actual burdening of religious practice of any magnitude is automatically deemed to be “an incidental effect.”

Hence, the Court’s holding in *Smith* is a doctrinal test for suppressing free exercise analysis, not for performing it. Because “few States would be so naive as to enact a law directly prohibiting or burdening a religious

40. *Id.*

41. *Id.* at 1600.

42. *Id.*

43. *Id.*

44. See *id.* at 1600-03.

45. *Id.* at 1608 (O’Connor, J., concurring).

practice as such,'⁴⁶ and thus meet the standard of a statute which targets and therefore burdens religious practice, only in an extraordinary circumstance does *Smith* authorize any individual assessment by a court of the actual magnitude of the claimed burdening of free exercise.⁴⁷

B. *The Smith Direct/Indirect Dichotomy*

Given the critical emphasis in the Court's holding and reasoning on the direct/indirect dichotomy,⁴⁸ there are two meanings that must be considered in decoding the Court's reference to "merely the incidental effect of a generally applicable and otherwise valid provision."⁴⁹ The first meaning is that if one looks from the vantage point of the legislative intent animating a facially neutral statute, then any impact in burdening free exercise may validly be called an "incidental effect" irrespective of the particular nature and magnitude of the burden. To illustrate, the primary legislative intent of the Oregon statute, or any statute criminalizing the possession of narcotics, is to discourage the use of drugs and to punish those who violate the statute. Thus, because an impact on free exercise is not intended by the legislature, any actual intrusion on free exercise may be characterized as an "incidental effect." In this first meaning, therefore, the words an "incidental effect" embrace free exercise claims from the most minor to the most egregious. All such

46. *Id.*

47. Justice O'Connor, in her concurring opinion, argued for the need for individual assessment of free exercise claims as a requirement of the judicial function:

To me, the sounder approach—the approach more consistent with our role as judges to decide each case on its individual merits—is to apply [the *Sherbert* compelling interest] test in each case to determine whether the burden on the specific plaintiffs before us is constitutionally significant and whether the particular criminal interest asserted by the State before us is compelling. Even if, as an empirical matter, a government's criminal laws might usually serve a compelling interest in health, safety, or public order, the First Amendment at least requires a case-by-case determination of the question, sensitive to the facts of each particular claim. Given the range of conduct that a State might legitimately make criminal, we cannot assume, merely because a law carries criminal sanctions and is generally applicable, that the First Amendment *never* requires the State to grant a limited exemption for religiously motivated conduct.

Id. at 1611 (citation omitted) (emphasis in original).

48. The Court's emphasis on "merely the incidental effect" invokes the direct/indirect dichotomy and a comment about the dichotomy in a different context: "In thus making use of the expressions, 'direct' and 'indirect interference' with commerce, we are doing little more than using labels to describe a result rather than any trustworthy formula by which it is reached." *Di Santo v. Pennsylvania*, 273 U.S. 34, 44 (1927) (Stone, J., dissenting).

49. *Employment Div. v. Smith*, 110 S. Ct. 1595, 1600 (1990).

unintended intrusions are fairly characterized as adding up only to an “incidental effect” even if there is an egregious burden on religious exercise.

In the second meaning of an “incidental effect,” however, there is a focus on the nature and magnitude of the actual burden inflicted. The Court’s words in *Thomas*, which are quoted in *Hobbie*, incorporate this distinction: “While the compulsion may be indirect, the infringement upon free exercise is nonetheless substantial.”⁵⁰ This meaning reflects the perspective of the person whose free exercise of religion is burdened and who surely will be concerned about the nature and magnitude of the intrusion. If the burdening of religious exercise is substantial, it is not “merely [an] incidental effect.”⁵¹ To collapse the two meanings into one obscures the second meaning. The *Smith* decision embraces only the first meaning. Because it bars any scrutiny of the specific facts, the *Smith* decision excludes any judicial assessment of the exact nature and magnitude of the intrusion on the free exercise of religion.

C. The Court’s Passionate Language and Suppression of Free Exercise Analysis

The Court’s emphatic intent to suppress free exercise analysis is further evidenced by its vivid and even passionate language. The Court explicitly rejected the compelling interest balancing test for performing free exercise analysis. It also repudiated the test in the strongest possible language: “[I]t is horrible to contemplate that federal judges will regularly balance against the importance of general laws the significance of religious practice.”⁵² An across-the-board application of the compelling interest

50. *Hobbie v. Unemployment Appeals Comm’n*, 480 U.S. 136, 141 (1987); *Thomas v. Review Bd.*, 450 U.S. 707, 718 (1981).

51. See *Smith*, 110 S. Ct. at 1600.

52. *Id.* at 1606 n.5. Central to the justification for the *Smith* majority’s categorical rule is the claim that application of the compelling interest test would require the courts to “constantly be in the business of determining whether the ‘severe impact’ of various laws on religious practice . . . suffices to permit us to confer an exemption.” *Id.* (emphasis added). Justice Scalia makes perfectly clear that it is the frequency of the balancing of “the significance of religious practice” against “the importance of general laws” that is deeply objectionable. *Id.* The core significance of this argument is also apparent from the Court’s emphatic language:

Moreover, if “compelling interest” really means what it says . . . many laws will not meet the test. Any society adopting such a system would be courting anarchy, but that danger increases in direct proportion to the society’s diversity of religious beliefs, and its determination to coerce or suppress none of them. Precisely because “we are a cosmopolitan nation made up of people of almost every conceivable religious preference,” and precisely because we value and protect that religious divergence, we cannot afford the luxury of deeming pre-

test would be “courting anarchy,” a luxury our “cosmopolitan nation

sumptively invalid, as applied to the religious objector, every regulation of conduct that does not protect an interest of the highest order. The rule respondents favor would open the prospect of constitutionally required religious exemptions from civic obligations of almost every conceivable kind — ranging from compulsory military service, to the payment of taxes, to health and safety regulation such as manslaughter and child neglect laws, compulsory vaccination laws, drug laws, and traffic laws, to social welfare legislation such as minimum wage laws, child labor laws, animal cruelty laws, environmental protection laws, and laws providing for equality of opportunity for the races. The First Amendment’s protection of religious liberty does not require this.

Id. at 1605-06 (citations omitted) (emphasis in original).

In her concurring opinion, Justice O’Connor replied: “The Court’s parade of horribles . . . not only fails as a reason for discarding the compelling interest test, it instead demonstrates just the opposite: that courts have been quite capable of applying our free exercise jurisprudence to strike sensible balances between religious liberty and competing state interests.” *Id.* at 1612-13. Justice Scalia responded, in a footnote, to this criticism by arguing that:

Justice O’Connor mistakes the purpose of our parade: it is not to suggest that courts would necessarily permit harmful exemptions from these laws (though they might), but to suggest that courts would *constantly* be in the business of determining whether the “severe impact” of various laws on religious practice . . . suffices to permit us to confer an exemption.

Id. at 1606 n.5 (emphasis added). The best guide for determining the frequency of free exercise claims is to look to the past. The *Sherbert* compelling interest test has existed since 1963. No floodgate was opened by *Sherbert* and its progeny. In Justice Blackmun’s words:

This Court, however, consistently has rejected similar arguments in past free exercise cases, and it should do so here as well. The State’s apprehension of a flood of other religious claims is purely speculative. Almost half the States, and the Federal Government, have maintained an exemption for religious peyote use for many years, and apparently have not found themselves overwhelmed by claims to other religious exemptions. Allowing an exemption for religious peyote use would not necessarily oblige the State to grant a similar exemption to other religious groups. The unusual circumstances that make the religious use of peyote compatible with the State’s interests in health and safety and in preventing drug trafficking would not apply to other religious claims. Some religions, for example, might not restrict drug use to a limited ceremonial context, as does the Native American Church. Some religious claims involve drugs such as marijuana and heroin, in which there is significant illegal traffic, with its attendant greed and violence, so that it would be difficult to grant a religious exemption without seriously compromising law enforcement efforts. . . . Though the State must treat all religions equally, and not favor one over another, this obligation is fulfilled by the uniform application of the “compelling interest” *test* to all free exercise claims, not by reaching uniform *results* as to all claims. A showing that religious peyote use does not unduly interfere with the State’s interests is “one that probably few other religious groups or sects could make.”

Id. at 1620-21 (Blackmun, J., dissenting) (citations omitted) (emphasis in original).

The wider specter of the floodgate argument (“many laws will not meet the test”)

. . . cannot afford,”⁵³ creating “a system in which each conscience is a law unto itself or in which judges weigh the social importance of all laws against the centrality of all religious beliefs”⁵⁴ and “produc[ing] . . . a private right to ignore generally applicable laws.”⁵⁵

is contradicted by the record of our modest free exercise case law. The challenged laws have already met the test in Justice Scalia’s “parade of horribles,” and the frequency of cases is not large. Indeed, the parade detailed in the *Smith* majority’s specification of 12 cases in the highest state and federal courts from 1941 through 1989 does not support his floodgate argument. *See id.* at 1605-06. The parade is far too short, and it supports Justice O’Connor’s argument that the courts can “strike sensible balances between religious liberty and competing state interests.” *Id.* at 1613. Without a concrete basis, the floodgate argument is pure speculation and apprehension. Finally, this genre of instrumentalist argument should be considered in light of the status of the “First Amendment’s command that religious liberty is an independent liberty, that it occupies a preferred position.” *Id.* at 1609 (O’Connor, J., concurring). For those jurists committed to this cardinal principle, the floodgate argument does not come easily here.

53. *Id.* at 1605.

54. *Id.* at 1606.

55. *Id.* at 1604. This passionate, even fearful, language may be related to the *Smith* majority’s repeated misstatement and dramatization of the rule urged upon the Court by Smith and Black: “Respondents urge us to hold, quite simply, that when otherwise prohibitable conduct is accompanied by religious convictions, not only the convictions but the conduct itself must be free from governmental regulation.” *Id.* at 1602. “They assert . . . that ‘prohibiting the free exercise [of religion]’ includes requiring any individual to observe a generally applicable law that requires (or forbids) the performance of an act that his religious belief forbids (or requires).” *Id.* at 1599. “The rule respondents favor would open the prospect of constitutionally required religious exemptions from civic obligations of almost every conceivable kind. . . .” *Id.* at 1605. In sum, the Court majority argues in classic either/or fashion that the only alternative to its sweeping categorical rule is “courting anarchy.” *See id.*

Actually, the respondents’ brief plainly indicated that their claim was utterly conventional in arguing in familiar positivist fashion that the *Sherbert* compelling interest balancing test, embedded in the Court’s four prior unemployment insurance precedents, should be applied to their unemployment insurance case: “[T]here is nothing to distinguish the present case from this Court’s prior unemployment decisions in *Sherbert*, *Thomas*, *Hobbie* and *Frazee*. Respondents urge the Court to affirm the decision of the Oregon Supreme Court . . . on this basis.” Brief for Respondents at Point IIA, Employment Div. v. Smith, 110 S. Ct. 1595 (1990). The Brief for Respondents does not argue for a new test or for a reinterpretation of the embedded *Sherbert* test. Indeed, the Court itself explicitly acknowledges that “[r]espondents’ claim for relief rests on our decisions in *Sherbert* v. *Verner*, *Thomas* v. *Review Board* . . . and *Hobbie* v. *Unemployment Appeals Comm’n*, in which we held that a State could not condition the availability of unemployment insurance on an individual’s willingness to forgo conduct required by his religion.” *Smith*, 110 S. Ct. at 1598.

The *Smith* Court’s hostility to the application of the free exercise clause may also be manifested in the Court’s reliance on the decision in *Minersville School District v. Gobitis*, 310 U.S. 586 (1940), without noting that *Gobitis* was explicitly overruled eight to one only three years later in *West Virginia State Board of Education v. Barnette*, 319

Given the Court's analysis in other first amendment areas,⁵⁶ the rejection of the compelling interest balancing test could easily have led to the Court's substitution of another, less challenging balancing test. In its prior free exercise cases, far less challenging tests are suggested, such as a "reasonable means" test and a weakened version of the compelling interest test.⁵⁷ Yet, the *Smith* Court's hostility to judicial enforcement of the free exercise clause is so consuming that even the "reasonable means" test, a veritable engine of justification that can validate almost any statute, is not authorized. The stark fact is that no test is substituted to assess a claimed burdening of religious practice. The reason is perfectly clear. The *Smith* holding and reasoning are patently designed to suppress, not authorize, free exercise analysis.

U.S. 624 (1943). Justice Scalia stated:

We have never held that an individual's religious beliefs excuse him from compliance with an otherwise valid law prohibiting conduct that the State is free to regulate. On the contrary, the record of more than a century of our free exercise jurisprudence contradicts that proposition. As described succinctly by Justice Frankfurter in *Minersville School Dist. Bd. of Educ. v. Gobitis*, 310 U.S. 586, 594-95, 60 S. Ct. 1010, 1012-13, 84 L. Ed. 1375 (1940): "Conscientious scruples have not, in the course of the long struggle for religious toleration, relieved the individual from obedience to a general law not aimed at the promotion or restriction of religious beliefs. The mere possession of religious convictions which contradict the relevant concerns of a political society does not relieve the citizen from the discharge of political responsibilities."

Smith, 110 S. Ct. at 1600 (footnote omitted). *Gobitis* is cited a second time on the same page, again without any indication that it was overruled three years later. *Id.* In addition, in concurring in a recent case, *Barnes v. Glen Theatre, Inc.*, 111 S. Ct. 2456 (1991), Justice Scalia again repeated the first sentence in the above quotation without any indication that *Gobitis* was overruled. In *Barnette*, Justice Jackson, speaking for the Court, upheld on general first amendment grounds the right of students who were Jehovah's Witnesses to refuse to salute the flag. The *Barnette* Court found it unnecessary "to inquire whether non-conformist beliefs will exempt from the duty to salute unless we first find power to make the salute a legal duty." *Barnette*, 319 U.S. at 635. In often quoted language, the Court stated: "If there is any fixed star in our constitutional constellation, it is that no official, high or petty, can prescribe what shall be orthodox in politics, nationalism, religion, or other matters of opinion or force citizens to confess by word or act their faith therein. If there are any circumstances which permit an exception, they do not now occur to us." *Id.* at 642. For broader insight into Justice Scalia's jurisprudence, including his use of inflammatory language, see Kannar, *The Constitutional Catechism of Antonin Scalia*, 99 YALE L.J. 1297 (1990).

56. See *Employment Div. v. Smith*, 110 S. Ct. 1595, 1612 (1990) (O'Connor, J., concurring)..

57. In *Bowen v. Roy*, 476 U.S. 693, 707-08 (1986), three justices suggested that it is sufficient if a facially neutral, generally applicable rule providing benefits is a "reasonable means of promoting a legitimate public interest." In *United States v. Lee*, 455 U.S. 252, 259 (1982), the compelling interest test that must be demonstrated is that the free exercise claim not "unduly interfere with fulfillment of the governmental interest." For a detailed critique of reasonableness as a first amendment test, see C.E. BAKER, HUMAN LIBERTY AND FREEDOM OF SPEECH 126-31 (1989).

D. *The Jurisprudence of Generally Applicable Criminal Statutes*

The *Smith* holding and principle, derived from a premise that attributes extraordinary power to an ordinary criminal statute as a bar to a free exercise claim, comprise the framework that the Court used to distinguish its prior case law, to reject the balancing test, and to justify its conclusions. Justice O'Connor called this attribution of power to the typical criminal statute “talismanic.”⁵⁸ Yet, paradoxically, this decisive, even talismanic, importance of the typical criminal statute is postulated, never elaborated, and indeed never even directly argued.⁵⁹

Any critique of the Court’s major premise in *Smith* must begin by unfolding what the Court calls the importance of the governmental interest in across-the-board criminal statutes: the jurisprudence of facially neutral, generally applicable criminal statutes, including their precise nature, scope, and purpose in our system of constitutional criminal law. This affirmative exposition also leads to an understanding of their limitations and to a critique of the Court’s simplistic and erroneous view of the governmental interest at stake.

In essence, these typical criminal statutes are designed to serve the overriding purposes of the criminal law within our constitutional system, including the protection of the fundamental interests embedded in the Bill of Rights and in the framework of divided, separated, and checked powers. This critique begins in Part II with an assessment of the importance of these statutes from the perspective of our theory of just punishment.

II. CRIMINAL STATUTES AND JUST PUNISHMENT

A. *Across-the-Board Prohibitions*

The *Smith* Court’s attribution of talismanic power to an ordinary across-the-board criminal statute to defeat any claim of burdening of free exercise violates our traditional theory of just punishment. Just punishment presupposes just liability. Just liability requires that those

58. *Smith*, 110 S. Ct. at 1612 (O’Connor, J., concurring).

59. The Court does, of course, argue intensely that considering the judicial creation of free exercise exemptions to typical criminal statutes will be “courting anarchy,” a luxury our “cosmopolitan nation . . . cannot afford.” *Id.* at 1605. This vivid rhetoric assumes what is disproved in Part II of this Article: that the criminal law does not have an existing system of exemptions and that “an unbending application of a criminal prohibition” is the norm so that creating free exercise exemptions risks “anarchy.” See *id.* at 1617 (Blackmun, J., dissenting). What is not argued, however, is why typical criminal statutes, within the history, theory, and practice of American criminal law, have the automatic and extraordinary power to bar a free exercise claim.

who demonstrate the same *mens rea* embodied in a criminal law act that causes a proscribed harm be treated equally in measuring liability. Thus, those who demonstrate an intent to kill embodied in a criminal act that causes death are liable for intentional murder.⁶⁰ Those who demonstrate an intent to rob and who take property by force or fear are liable for robbery.⁶¹ Those who take the personal property of another with the intent to steal are liable for larceny.⁶² Those who demonstrate an intent to cause physical injury to another person and who cause such injury are liable for assault.⁶³ And so on.

These facially neutral across-the-board criminal prohibitions serve to calibrate moral fault by providing for the channeling of like cases into categories for like treatment. They therefore serve the ideal of equal justice required for just liability. This channeling of like cases serves a closely related requirement for retributive justice. Different crime categories provide different legislative measures of moral fault. For example, robbery, a larceny with force or threat of force, exhibits more serious moral fault than simple larceny and therefore, a more severe penalty is authorized. In addition, the common degree structure that exists within each major crime category reflects legislative refinements of moral fault. Robbery with a dangerous weapon is typically robbery in the first degree with a more severe scope of penalties, yet robbery without such a weapon is robbery in the second or third degree with a lesser scope of penalties.⁶⁴

B. Across-the-Board Defenses

Our theory of just liability also requires that those who demonstrate an intent to kill embodied in a criminal law act that causes death are not liable for intentional murder if they are insane.⁶⁵ In other words, they are liable only for manslaughter if they kill in response to legally recognized provocation.⁶⁶ Those who rob another are not liable for robbery if they are legally infants.⁶⁷ Those who steal personal property are not liable for larceny if they are the victim of duress or justification.⁶⁸ Those who strike another with the intent to cause physical injury are

60. See, e.g., N.Y. PENAL LAW § 125.25 (McKinney 1987); MODEL PENAL CODE § 210.2 (1962).

61. See, e.g., N.Y. PENAL LAW § 160.00; MODEL PENAL CODE § 222.

62. See, e.g., N.Y. PENAL LAW § 155.05; MODEL PENAL CODE § 223.2.

63. See, e.g., N.Y. PENAL LAW § 120.00; MODEL PENAL CODE § 211.1.

64. See, e.g., N.Y. PENAL LAW §§ 160.05, 160.15; MODEL PENAL CODE § 222.1.

65. See, e.g., N.Y. PENAL LAW § 40.15; MODEL PENAL CODE § 4.01.

66. See, e.g., N.Y. PENAL LAW § 35.00; MODEL PENAL CODE § 210.3.

67. See, e.g., N.Y. PENAL LAW § 30.00; MODEL PENAL CODE § 4.10.

68. See, e.g., N.Y. PENAL CODE §§ 35.05(2), 40.00 (duress and justification); MODEL PENAL CODE §§ 2.09, 3.01 (duress and justification).

not liable for assault if they strike in self-defense or defense of another.⁶⁹ And so on, across the range of traditional defenses.

Just liability also means that these facially neutral across-the-board statutory defenses are equally essential in calibrating moral fault by the test of like treatment of like cases and by the test of retributive justice. Our theory of just liability, rooted in the principles of equal justice and retributive justice, does not authorize the grouping together of the sane and insane, adults and children, those who have the capacity for free choice and those who do not, and those who attack others without justification or excuse and those who strike others in self-defense or defense of another.

To have just liability there must be sufficient moral fault, and there is either no moral fault or diminished moral fault if a justification or excuse exists. For just liability, there must also be equal treatment of those who are similarly situated. Those whose behavior falls into a recognized category of justification or excuse are not similarly situated with those whose behavior must be categorized differently. To illustrate, if A shoots and kills B in a robbery and C shoots and kills D in self-defense or defense of another, the behavior of killing by shooting is identical, but A is morally culpable and liable for murder while C's behavior is justified. C is morally innocent and not legally blameworthy. Although the external behavior is identical, A and C fit into radically different moral and legal categories.⁷⁰ In determining whether fault exists,

69. See, e.g., N.Y. PENAL LAW § 35.15; MODEL PENAL CODE §§ 3.04, 3.05 (self-protection and protection of others).

70. This moral calibration is demonstrated by the structure of our homicide law. The modern structure of homicide categories is an intricate web grounded in hundreds of years of common-law development. See G. FLETCHER, RETHINKING CRIMINAL LAW, 235-40 (1978). The homicide web of categories details subtle degrees of moral fault as measured by closely related but distinct forms of *mens rea*. Those who kill with an intent to kill (*i.e.*, a conscious design to kill, along with deliberation and premeditation) are grouped together in some states as liable for murder in the first degree, the most egregious category of murder. Those who kill with an intent to kill unaccompanied by premeditation and deliberation are categorized separately in some states as liable for murder in the second degree. The different degrees of moral fault are matched with different types of severe punishment, including life imprisonment and death. Those who have an intent to inflict serious bodily injury are grouped together as liable for a lesser form of murder or manslaughter. Those who kill with an intent to kill triggered by heat of passion or in response to certain types of recognized provocation or extreme emotional disturbance are grouped together in a manslaughter category. Those who kill without an intent to kill, but with a depraved indifference to human life are grouped together in a murder category and are distinguished from those who kill with a reckless form of moral fault (reckless manslaughter). Both groups are distinguished from those who kill due to criminal negligence, a killing with inherently less moral fault than one committed recklessly or with intent. *Id.*

the unit for analysis is the behavior of shooting and killing another and the *reason* for that behavior.

The ethos of equal treatment, as well as retributive justice, compel these distinctions, which are exemplified in the panorama of criminal law defenses rooted in our religious values, our culture, and hundreds of years of common-law tradition. The rules concerning justification and excuse were first systematically developed in the twelfth century by Abelard and the canonists.⁷¹ In modern form, these distinctions are also generally codified in our penal laws and routinely applied every day in many thousands of cases.⁷² In ancient times, Aristotle articulated the conflict between the need in law "to speak universally" and defined equity as the "correction of the law where it is defective by reason of its universality."⁷³

C. *The Need for Balancing*

The calibration of moral fault sufficient to determine just liability compels a balancing of two interests: the interests protected by the criminal prohibitions such as murder, manslaughter, assault, robbery, and larceny and the interests protected by the defenses and partial defenses such as self-defense, defense of another, extreme emotional disturbance, entrapment, and insanity. The existence of facially neutral across-the-board criminal prohibitions does not bar defenses that are also facially neutral and across-the-board, nor do such prohibitions estop balancing or foreclose analysis of such claims.

This balancing is not simply at the level of facial analysis of statutes that define crimes and defenses. The finding of just liability by balancing

71. H. BERMAN, LAW AND REVOLUTION, THE FORMATION OF THE WESTERN LEGAL TRADITION 187-90 (1983).

72. The *Smith* case also illustrates this analysis. The criminal prohibition, which the *Smith* majority found decisive, is the facially neutral across-the-board Oregon criminal prohibition of the knowing or intentional possession of a controlled substance unless the substance has been prescribed by a medical practitioner. See OR. REV. STAT. § 475.992(4) (1987). Yet, a person is patently not liable under this statute, even when there is proof of these elements beyond a reasonable doubt, if the person is a victim of duress, an infant, or is insane. See OR. REV. STAT. §§ 161.270, 161.290, 161.295 (1990). A shorthand explanation for the lack of liability is that the person who can establish any one of these defenses lacks moral fault and thus lacks the *mens rea* required for just liability. In addition, even though moral guilt may exist, the person may not be liable if the proof of criminal guilt did not meet the constitutional proof standard of beyond a reasonable doubt or the drugs were seized in violation of the fourth amendment prohibition of unreasonable searches and seizures. See *In re Winship*, 397 U.S. 358 (1970); *Mapp v. Ohio*, 367 U.S. 643 (1961). Our theory of just punishment requires that across-the-board prohibitions against particular behaviors be viewed in light of underlying requirements of the individual's capacity, constitutional principles, and other relevant defenses.

73. See H. BERMAN, *supra* note 71, at 518.

competing interests also requires the application of the principles of equal and retributive justice in the individual case in light of its distinctive facts. In addition to justifying a statutory prohibition on its face, the application of a statute must also be justified in light of the distinctive facts of the case. Facial neutrality, in the sense of equal and retributive justice, does not automatically result in the vindication of these principles in application. The most pristine statute, from the standpoint of justice, may be enforced unjustly. Case-by-case scrutiny is essential to assure that the statute's promise of justice is made real in particular cases. Conscientious police, prosecutors, judges, and defense lawyers struggle in every case to make these principles real.

Hundreds of years of experience demonstrate that not everyone who initially appears to fit neatly into a generally applicable charge category deserves to be placed into that category upon scrutiny of the specific facts of a case. Indeed, it is precisely the role of the prosecutor and the defense lawyer to present evidence and arguments concerning which category should apply. In addition, it is the province of judge and jury to decide whether the behavior means that a particular defendant fits into the category created by a "generally applicable criminal prohibition" or is "exempted" from such a category because the facts mandate that the defendant be placed in the category created by a generally applicable defense.⁷⁴

D. "*Without Good Cause*" and *Mens Rea*

In both unemployment compensation and criminal law, the blend of external behavior and the reason for the behavior comprises the micro cosmos that must be scrutinized for its legal meaning. The unemployment compensation tests of misconduct or "without good cause"⁷⁵ for quitting or refusing a job are matched by the criminal law test of *mens rea*, the measure of moral fault required for criminal liability. Like the "good cause" test in the unemployment compensation area, the *mens rea* requirement for just liability mandates a criminal law "system of in-

74. Trial language and procedure illustrate the balancing of interests required to realize individual justice. A judge, to make good on her oath to uphold the laws, must apply all the relevant laws pertaining to both the charges and defenses at the hearing or trial. Once the defense meets its burden of production for raising a defense, the defense is cognizable and must be adjudicated according to ordinary trial procedure. For example, the prosecutor has the burden of proving guilt by the difficult standard of proof beyond a reasonable doubt, so that usually a defense prevails if only a reasonable doubt as to guilt is raised. The judge must carefully define both the relevant charging statutes and defenses to guide the jury's deliberation, and the jury must then decide which category applies, charge or defense.

75. See *Employment Div. v. Smith*, 110 S. Ct. 1595, 1603 (1990).

dividual exemptions,"⁷⁶ which must be applied in each case as the particular facts require.

In deciding each case, behavior must be assessed. In unemployment compensation cases, this behavior includes quitting work or refusing to accept available work.⁷⁷ In criminal law, the behavior prohibited by the specific statute, the *actus reus*, must be assessed.⁷⁸ In each case, statutory tests are provided by which to assess the reasons offered for the behavior. In unemployment compensation, the "without good cause" or "misconduct" standard applies to assess the behavior of quitting work or refusing available work.⁷⁹ As the *Smith* Court emphasized, the application of this standard "invite[s] consideration of the particular circumstances behind an applicant's unemployment" and creates "a mechanism for individualized exemptions."⁸⁰

In criminal law, the core principle of *mens rea*, as detailed in the range of related defenses, applies to test the external behavior prohibited by the statute. If the behavior prohibited was committed in self-defense or in the defense of another, that is, a "good cause" for the behavior, the person is not liable. In criminal law language, the act that harms another is justified because the defendant has a reason that we recognize as exculpatory. The defendant is not blameworthy because she lacked the required fault expressed in the *mens rea* element. This exemption from liability is many hundreds of years old. Indeed, by the twelfth century, "[i]t was accepted that a person who intentionally attacks another may be justified by self-defense or by defense of others."⁸¹

In addition, the independent principle of *actus reus* requires that the criminal law act itself be voluntary, "an exercise of free choice externalized into overt behavior."⁸² If it is involuntary, as in a reflex or convulsion, no *actus reus*, as well as no *mens rea*, exists and hence,

76. *Id.*

77. See *Frazee v. Illinois Dep't of Employment Sec.*, 489 U.S. 829, 830 (1989); *Hobbie v. Unemployment Appeals Comm'n*, 480 U.S. 136, 137 (1987); *Thomas v. Review Bd.*, 450 U.S. 707, 709 (1981); *Sherbert v. Verner*, 374 U.S. 398, 399 (1963).

78. See, e.g., N.Y. PENAL CODE § 15.10 (McKinney 1987); MODEL PENAL CODE § 2.01 (1962).

79. See *Smith*, 110 S. Ct. at 1603.

80. *Id.* (quoting *Bowen v. Roy*, 476 U.S. 693, 708 (1986)). The definition of "without cause" or "misconduct" can incorporate the tone and even some of the traditional symbols of *mens rea*. For example, Florida's statute, which was reviewed in *Hobbie*, defines "misconduct" in part as: "(b) Carelessness or negligence of such a degree or recurrence as to *manifest culpability, wrongful intent, or evil design* or to show an intentional and substantial disregard of the employer's interests or of the employee's duties and obligations to his employer." *Hobbie*, 480 U.S. at 139 n.3 (quoting FLA. STAT. § 443.036(24) (1985)).

81. H. BERMAN, *supra* note 71, at 190.

82. J. DELANEY, CRIMINAL LAW: A PROBLEM SOLVING APPROACH 102 (1986).

no liability is present on either ground.⁸³ Even apart from *mens rea*, therefore, the finding of just liability mandates that the behavior itself be subject to scrutiny to verify that it is the expression of an incriminating reason and not simply an uncontrollable striking out which injures another person. In each criminal case, therefore, the foundational principles of *mens rea* and *actus reus* require, not simply invite, a “consideration of the particular circumstances behind” a person’s behavior, and the resulting scope of defenses create a centuries-old “mechanism for individualized exemptions.”⁸⁴ Indeed, “the canonists of the late eleventh and the twelfth centuries founded their doctrines of the subjective and objective aspects of crime . . . [on] ‘the precise investigation in any given case of the intention . . . and of the external circumstances of the act.’”⁸⁵ Though initially created at common law, the modern scope of justification and excuse is refined and codified in penal codes in each state and also detailed in the Model Penal Code authored by the American Law Institute.⁸⁶

Thus, it is plain that “a distinctive feature” of both our system of unemployment compensation and our system of criminal law is the requirement of “individualized governmental assessment of the reasons for the relevant conduct.”⁸⁷ There are reasons for behavior that our culture and our civil and criminal laws validate and reasons that our culture and our laws invalidate. Each case must be scrutinized individually to determine which reasons underlie the behavior and which are valid and which are invalid. In both systems, a simplistic behavioristic preoccupation with the external conduct alone, a “snapshot” mirroring overt behavior, is insufficient. Such a preoccupation reduces what is at stake. Justice requires that those who make factual and legal findings, judge and jury alike, consider both the external behavior and the reasons for the behavior. It is precisely the role of judge and jury to determine this meaning by applying across-the-board charging and defense statutes in individual cases to determine whether a particular behavior fits into one legal category rather than the other. Thus, the principle that the Court extracts from the unemployment compensation cases, *Sherbert*, *Thomas*, *Hobbie*, and *Frazee*, is applicable to the criminal law arena. As the Court stated, “[O]ur . . . cases stand for the proposition that where the State has in place a system of individual exemptions, it may

83. *Id.* at 102-04.

84. See *Employment Div. v. Smith*, 110 S. Ct. 1595, 1603 (1990) (quoting *Bowen v. Roy*, 476 U.S. 693, 708 (1986)).

85. H. BERMAN, *supra* note 71, at 192 (quoting G. LE BRAS, “CANON LAW,” THE LEGACY OF THE MIDDLE AGES 357 (1926)).

86. MODEL PENAL CODE §§ 1.01-405.4 (1962).

87. *Smith*, 110 S. Ct. at 1603.

not refuse to extend that system to cases of ‘religious hardship’ without compelling reason.”⁸⁸

Though the rationale expressed in the Court’s unemployment cases easily applies, the *Smith* Court, nevertheless, emphatically rejected the application of the four decisions embodying the *Sherbert* compelling interest balancing test to the criminal context by stating, “[these cases] have nothing to do with an across-the-board criminal prohibition on a particular form of conduct.”⁸⁹ Thus, under *Smith*, the classic role of the criminal court to do justice by applying general statutes in individual cases to find that a person’s conduct fits into one category rather than another, has an ironic exception when the defendant’s claim is a burdening of free exercise of religion. In an additional irony, the *Smith* majority held that the assertion of this fundamental claim, which usually triggers individualized “as applied” analysis and “strict scrutiny” through the compelling interest test, receives no scrutiny at all if it arises from the application of a typical criminal statute.

The rationale for the Court’s refusal to apply the compelling interest test in *Smith* is rooted in its false polarity of unemployment compensation as creating a system of “individualized governmental assessment” of the reasons for quitting or refusing work, as compared to “an across-the-board criminal prohibition of a particular form of conduct” without such a system of individualized exemptions.⁹⁰ As the above analysis indicates, however, our theory of a just criminal law has for centuries included a system of individualized exemptions. The claimed polarity is clearly erroneous and is oblivious of our criminal law history and jurisprudence.

E. The Distinction Collapses

Once this underlying rationale falls, the Court’s distinction collapses. Why should facially neutral unemployment compensation statutes be assessed regarding their application to the particular facts presented in each case while facially neutral criminal statutes require no such assessment? Criminal law jurisprudence, as applied to charging statutes, also supports this critique. Although such statutes appear to be complete definitions capable of discrete application (*e.g.*, robbery, arson, larceny), this atomistic conception is incorrect and misleading. To the contrary, they are, in Jerome Hall’s words, “neither autonomous nor complete.”⁹¹

88. *Id.*

89. *Id.*

90. *Id.*

91. J. HALL, GENERAL PRINCIPLES OF CRIMINAL LAW 19 (1960).

All such statutory definitions are incomplete and must be seen as part of a web of other statutes that define the across-the-board defenses relevant in a particular case. Hence, as Hall points out, “the [charging] rules do not completely define the specific crimes. . . [O]nly after the [defense] doctrines have been added to the rules has the penal law, i.e. the definitions of all the specific crimes, been fully stated.”⁹² To illustrate, when a person is charged with arson of a home, the charge explicitly and impliedly states, “one who sets fire to a dwelling-house, not being insane, intoxicated, an infant, coerced, etc., commits arson . . . [or] stated affirmatively: a sane, sober adult, acting freely . . . who sets fire.”⁹³

Hence, our criminal law jurisprudence mandates that the definitions specified in criminal law prohibitions require a system of individualized exemptions from liability for the behavior specified in across-the-board prohibitions. In addition, it is impossible to apply this system of individualized exemptions to particular cases without scrutinizing the specific facts posed in each case so that the relevant across-the-board defenses can be applied and adjudicated. The scrutiny of application is therefore a jurisprudential *sine qua non*. The *Smith* Court’s prohibition of any scrutiny of application beyond facial analysis of the statute bars the pursuit of individualized justice when a free exercise claim is presented.

F. More Important in Criminal Law

There is an independent reason, grounded in the centuries-old struggle to forge the core policy of preservation of liberty into our criminal law jurisprudence, for more carefully scrutinizing free exercise claims that arise in a criminal law context than in a civil law context. The liberty interest at stake is more fundamental and thus, the array of individual exemptions that may be raised as defenses is infused with a different significance than the array of “good cause[s]” that may be raised in the unemployment compensation area.⁹⁴

The *Smith* Court reaffirmed the *Sherbert* compelling interest balancing test to protect free exercise interests in the unemployment compensation realm when a significant monetary benefit is at stake, although it repudiated the applicability of this compelling interest test, or any real test at all, in the criminal law realm when liberty is at stake. Although monetary benefits are clearly important, in our tradition of constitutional criminal law, the free exercise liberty interest requires the most stringent

92. *Id.*

93. *Id.*

94. See *Employment Div. v. Smith*, 110 S. Ct. 1595, 1603 (1990).

protection. This difference was expressed by Justice O'Connor in her concurring opinion:

A State that makes criminal an individual's religiously motivated conduct burdens that individual's free exercise of religion in the severest manner possible, for it "results in the choice to the individual of either abandoning his religious principle or facing criminal prosecution." [A] neutral criminal law prohibiting conduct that a State may legitimately regulate is, if anything, *more* burdensome than a neutral civil statute placing legitimate conditions on the award of a state benefit.⁹⁵

In our jurisprudence, the threat to liberty and autonomy inherent in a criminal charge triggers a heightened scrutiny both of the criminal statute at issue as well as the application of the statute to the specific facts detailed in the state's prosecution. In addition to the range of traditional defenses to criminal charges, the panoply of fundamental constitutional protections apply in scrutinizing the prosecutor's evidence, including the presumption of innocence, the probable cause standard, the test of proof beyond a reasonable doubt, and the guarantees of the fourth, fifth, sixth, and eighth amendments.⁹⁶

In sharp contrast, these protections are inapplicable in an unemployment compensation case because liberty is not at stake. Yet, the *Smith* Court retains the *Sherbert* compelling interest test in assessing a free exercise claim to protect the applicant's interest in the monetary benefit provided by unemployment compensation, but rejects any real test to assess such a claim in the criminal law area when liberty is at stake and the "system of individualized exemptions" is even more important. The choice made by the *Smith* Court repudiates the foundational principle that a challenge to a liberty interest demands heightened protection, not lesser protection, or as in *Smith*, no real protection of a free exercise claim.

G. *Smith and Police Power Myopia*

This Article presupposes the validity of the Court's core framework that "our decisions . . . stand for the proposition that where the State has in place a system of individual exemptions, it may not refuse to extend that system to cases of 'religious hardship' without compelling

95. *Id.* at 1610-11 (O'Connor, J., concurring) (emphasis in original).

96. See, e.g., *United States v. Watson*, 423 U.S. 411 (1976); *In re Winship*, 397 U.S. 358 (1970); *Malloy v. Hogan*, 378 U.S. 1 (1964); *Gideon v. Wainwright*, 372 U.S. 335 (1963); *Mapp v. Ohio*, 367 U.S. 643 (1961); *Weems v. United States*, 217 U.S. 349 (1910).

reasons.”⁹⁷ Thus, this critique has been addressed to the Court’s failure to apply its principle to facially neutral criminal statutes. Yet, this central framework demonstrates that the *Smith* majority looked upon the free exercise clause with a telescope whose prism distorts the significance of the police power perspective and portrays a peculiar framework of separation of powers and federalism.

The *Smith* “proposition” renders enforcement of the free exercise clause hostage to the choice of a state legislature to create “a system of individual exemptions” whenever it enacts a generally applicable police power statute. If the legislature includes such a system in enacting statutes, the free exercise clause is operable, and a claim that free exercise is burdened will be cognizable by the courts. Yet, if the legislature rejects such a system in enacting its facially neutral statute, the free exercise claim is not operable, and any claim of burdening, however egregious, will not be cognizable by the courts. Enforcement of the first amendment “command that religious liberty is an independent liberty, that it occupies a preferred position,”⁹⁸ is contingent on the policies and political choices of state legislatures in enacting statutes. State legislatures definitively decide whether a free exercise claim arising from the application of a statute will be cognizable by the state and federal courts.⁹⁹

The *Smith* Court’s police power myopia also led to its use of the classical fallacy known as “denying the antecedent of a conditional statement.”¹⁰⁰ Stated simply, “[t]he proposition that ‘A implies B’ is not the equivalent of ‘non-A implies non-B,’ and neither proposition follows logically from the other.”¹⁰¹ Initially, the *Smith* “proposition”

97. *Smith*, 110 S. Ct. at 1603.

98. *Id.* at 1609 (O’Connor, J., concurring).

99. The comments by Pastor Richard John Neuhaus, a conservative Lutheran theologian, about *Smith* are relevant in assessing the legislative hegemony over judicial cognizance of free exercise claims. He stated:

But I would say that I am less alarmed about a parade of horribles, about the terrible things that are going to happen as a consequence of *Smith*, than I am — and I think other people involved in the churches and synagogues are — about what this says about whether in fact we believe that rights are established by nature and God as the Declaration says, whether we believe in the Pledge of Allegiance that this is a nation under God, or is this all just pious fluff?

Firing Line: An Extraordinary Supreme Court Decision (PBS television broadcast, July 24, 1990) (transcript available through Southern Educational Communication Association, P.O. Box 5966, Columbia, S.C. 29250). “The founding vision of the First Amendment religion clause, with its two parts, no establishment and free exercise, was . . . the founders’ intention to say that the state is limited by a higher sovereignty, and that religion in society . . . is the carrier of that witness.” *Id.*

100. *French v. State*, 266 Ind. 276, 291 n.1, 362 N.E.2d 834, 843 n.1 (1977) (DeBruler, J., dissenting).

101. *Id.*

is a valid hypothetical syllogism illustrating that “A implies A” or, as also formulated, if A then B. Its first part, “where the State has in place a system of individual exemptions,”¹⁰² is called the antecedent A. The remainder of the proposition is the consequent B: “it may not refuse to extend that system to cases of ‘religious hardship’ without compelling reason.”¹⁰³ This valid hypothetical syllogism uses familiar if-then reasoning and goes “to the heart of the Socratic method.”¹⁰⁴

Yet, the Court’s immediate conclusion rejecting the existence of a system of exemptions in the criminal context (non-A) led it to conclude that the compelling interest test is inapplicable (non-B). This invalid leap from non-A to non-B illustrates the fallacy of “denying the antecedent of a conditional statement.”¹⁰⁵ The Court’s consequent statement that “[w]e . . . hold the [compelling interest] test inapplicable”¹⁰⁶ (non-B), simply does not follow from the denial of the antecedent (non-A), unless A is the *sole* possible antecedent for B. Viewing antecedents as causes, a causal relation exists between A and B (A causes B), but A is *not* the only possible cause of B (non-As can also cause B).¹⁰⁷ The Court’s reasoning is a classic *non sequitur*, unless it posits the decision of state legislators as the *only* conceivable cause for the original consequent (*i.e.*, the application of the compelling interest test to cases of religious hardship to protect free exercise interests and values). This view, however, ignores that the command of the free exercise text itself is an independent cause.

If one adds the Court’s holding to its proposition and conclusion, the Court is not only saying that non-A (the absence of a system of exemptions) equals non-B (no compelling interest test), but it is also saying that non-A equals what might be called non-T, no balancing test of any kind. Recall that the Court’s rejection of the compelling interest test in the criminal context also means the rejection of any substitute test. Once the burdening of free exercise is found to arise from a typical criminal statute, it is deemed to be “merely the incidental effect” of the application of the statute and “the First Amendment has not been offended.”¹⁰⁸ Thus, non-A equals non-B and non-T. The policy and

102. Employment Div. v. Smith, 110 S. Ct. 1595, 1603 (1990).

103. *Id.*

104. R. ALDISERT, LOGIC FOR LAWYERS 159 (1989).

105. *Id.* at 162.

106. *Smith*, 110 S. Ct. at 1603.

107. From a psychological perspective, “people tend to overassess [the] degree of correlation between conditions because they consider only instances in which condition A in fact occurs in tandem with condition B, while ignoring cases in which condition A is not associated with condition B.” Margulies, “Who Are You To Tell Me That?: Attorney-Client Deliberation Regarding Nonlegal Issues and the Interests of Nonclients”, 68 N.C.L. REV. 213, 236 n.84 (1990).

108. Employment Div. v. Smith, 110 S. Ct. 1595, 1600 (1990).

political choices of state legislatures in including or omitting “a system of individual exemptions”¹⁰⁹ in enacting statutes (A or non-A) is decisive in determining whether the compelling interest test will be applied and whether a balancing test of any kind will be applied.

Why should the interests and values codified in the ordinary criminal law defenses be routinely cognizable and adjudicable while the fundamental interests and values embodied in the free exercise clause of the Bill of Rights are not cognizable and adjudicable? Why should the interests and values protected by general criminal prohibitions be routinely balanced in thousands of hearings and criminal trials throughout the country against the traditional array of defenses never be balanced against a free exercise claim unless the statute directly targets religious practice? Why, in the unemployment compensation area, where money is at stake, should the “system of individual exemptions” implementing the test of “good cause” justify free exercise analysis under the compelling interest test while the centuries-old criminal law system of individual exemptions, mandated by the need for fault before liberty is deprived, never authorize any free exercise analysis unless the statute directly targets religious practice? Why should state legislative choice in omitting such a system of individual exemptions in enacting statutes have the power to bar permanently any consideration of a free exercise claim?

Our theory of just liability mandates that across-the-board criminal law prohibitions be viewed in light of a range of across-the-board criminal law defenses. It is unjust to single out the free exercise clause and bar this constitutional claim from consideration merely because it arises from the application of a facially neutral, generally applicable criminal statute. Although they are a prerequisite for equal and retributive justice, such statutes do not have the “talismanic” power to trump the quest for justice in individual cases by expunging the free exercise reason from our jurisprudential sensibility.

III. CRIMINAL STATUTES AND CONSTITUTIONAL CRIMINAL LAW

A. *Respecting the Majority Voice*

The *Smith* Court’s attribution of talismanic power to a typical criminal statute also finds no justification in our system of constitutional criminal law. Part III of this analysis begins with an affirmative presentation of the nature, scope, and purpose of criminal statutes within our constitutional tradition. This analysis leads to a constitutional critique of the *Smith* Court’s attribution of extraordinary power to a typical

109. *Id.* at 1603.

criminal prohibition so that the free exercise clause is drained of meaning in the criminal context and indeed, even removed from analysis and decisionmaking.

Within our constitutional framework, criminal statutes implement the legislative form of sovereign power, commonly called police power, to enact laws for the public health, safety, and general welfare.¹¹⁰ The exercise of the police power includes an expression of the political role of the criminal law in our democracy. To illustrate, the revival and imposition of capital punishment corresponds to a strong public consensus in many states in favor of capital punishment. In addition, the “white-hot conflict between pro-life and pro-choice advocates is [both a constitutional and] a political struggle over whether abortion should be criminalized once again.”¹¹¹ The political character of the criminal law is also dramatically revealed in the heightening of penalties for drug, gun, and sex offenses and other violent crimes and in the “influential role of criminal-law issues in political campaigns ranging from the state legislature to the presidency.”¹¹² In less dramatic form, the core prohibitions of murder, robbery, assault, rape, burglary, and larceny have the strongest possible support of the electorate, so that political controversy about such core crimes tends to be concentrated on the appropriateness of the punishment for their violation.¹¹³

In our democratic society, the majoritarian voice is rightly heard and respected by our legislative representatives in their decisions to criminalize and decriminalize, to define the elements of crimes and defenses, and to heighten or lower penalties. Therefore, an important governmental interest exists in typical facially neutral, across-the-board criminal statutes that embody the dominant political consciousness. These statutes serve to implement the legislative role in our governmental system which separates powers among the legislature, the executive, and the judiciary. The executive role is implemented by prosecutors who, unlike legislators, have the power to apply these statutes to individual cases by initiating criminal charges. The judicial role is carried out by judges and jurors who, unlike legislators and prosecutors, have the power to decide guilt or innocence in individual cases. Criminal prohibitions are also essential to fulfill the ideal of the rule of law. Legislatures enact facially neutral, across-the-board criminal statutes which guide and check the power of police, prosecutors, judges, and jurors.¹¹⁴

110. See, e.g., *Barnes v. Glen Theatre, Inc.*, 111 S. Ct. 2456, 2462 (1991); *Day-Brite Lighting, Inc. v. Missouri*, 342 U.S. 421, 423 (1952).

111. J. DELANEY, *supra* note 82, at 12.

112. *Id.*

113. See *id.*

114. See, e.g., *Papachristou v. City of Jacksonville*, 405 U.S. 156, 165-71 (1972); *Grayned v. City of Rockford*, 408 U.S. 104, 108-09 (1972).

The Oregon statute utilized by the *Smith* majority, which prohibits the knowing and intentional possession of a controlled substance, clearly falls within the legislative exercise of the police power (*i.e.*, within the authorization of legislation for the public health, safety, and general welfare).¹¹⁵ Indeed, in *Robinson v. California*,¹¹⁶ the Supreme Court, in dictum, specifically authorized the prohibition of the possession of narcotics to be within the police power and found a California statute penalizing the “status” of being a drug addict to be outside of the police power.¹¹⁷ The Oregon possession statute illustrates, therefore, the importance in our democracy of what the Court in *Smith* calls facially neutral, across-the-board criminal prohibitions, because they implement both the rule of law and the democratic will of the majority.

B. Respecting the Bill of Rights

Although democratically determined police power embodies a basic and pervasive governmental interest, it is not unlimited. In our tradition of democratic criminal law, police power and the political struggle over its meaning is limited by federal and state constitutions. Indeed, the American contribution to political theory exemplified in our Constitution that protection of individual rights is central, not peripheral, to democracy reverberates throughout our criminal law.¹¹⁸ “Since democratic theory stresses both majority rule and protection of individual rights, there is an inherent tension between the political role of the criminal law, which implements shifting majoritarian priorities, and the constitutional role of the criminal law, which safeguards personal freedom and autonomy against even majoritarian will.”¹¹⁹ In our system of divided, separated, and checked power, this political role is carried out by the legislature,¹²⁰ while the safeguarding of personal freedom and autonomy against even a majoritarian will is the special, but not exclusive, responsibility of the judiciary.¹²¹

115. See OR. REV. STAT. § 475.992(4) (1987).

116. 370 U.S. 660 (1962).

117. *Id.* at 666.

118. In Ronald Dworkin’s words, “America’s principal contribution to political theory is a conception of democracy according to which the protection of individual rights is a pre-condition, not a compromise, of that form of government.” Dworkin, *The Reagan Revolution and the Supreme Court*, N.Y. Rev. of Books, July 18, 1991, at 23.

119. J. DELANEY, *supra* note 82, at 45.

120. L. TRIBE, *supra* note 7, at 19-22.

121. Thus, Madison, in presenting the proposed Bill of Rights to the Congress, stated:

If they are incorporated into the Constitution, independent tribunals of justice will consider themselves in a peculiar manner the guardians of those rights; they

Thus, criminal statutes and their enforcement must not infringe upon the fundamental interests embodied in the Bill of Rights, our charter of liberties. Those preferred rights create a framework of fundamental principles, a fence within which the police power must be exercised. For example, the fourteenth amendment prohibits introduction of even crucial state prosecutorial evidence if the police violate the fourth amendment ban on unlawful searches and seizures,¹²² the fifth amendment prohibition of testimonial compulsion,¹²³ or the sixth amendment guarantee of the right to counsel,¹²⁴ or if a statute violates the first amendment guarantees of freedom of expression.¹²⁵

The police, prosecutors, judges, and jurors, who are sworn to apply the legislature's criminal statutes to specific cases, are also sworn to apply them within the framework of our Bill of Rights. Even overwhelming evidence proving beyond all doubt that the defendant violated a criminal statute may not result in criminal liability if the government violates one of the historic guarantees that distinguish our culture from authoritarian and totalitarian societies. Notwithstanding even overwhelming evidence against the defendant, such governmental intrusion beyond its power may be sufficient reason to pluck the defendant from the category of culpability created by the charging statute and classify her within one of the categories of constitutional defenses created by these guarantees.

C. *Respecting the Twin Imperatives: Implementing Tests*

The fundamental police power and the core interests protected by the Bill of Rights exist in both a theoretical and an existential tension.¹²⁶

will be an impenetrable bulwark against every assumption of power in the Legislature or Executive; they will be naturally led to resist every encroachment upon rights expressly stipulated for in the Constitution by the declaration of rights.

Frantz, *The First Amendment in the Balance*, 71 YALE L.J. 1424, 1448 n.100 (1962) (quoting 1 ANNALS OF CONGRESS 139 (1789)). Yet, in emphasizing the role of the courts as a special guardian of the Bill of Rights, one must not overlook the obligation of all public officials, legislators, mayors, governors, and presidents to respect the Bill of Rights.

122. See *Mapp v. Ohio*, 367 U.S. 643 (1961).

123. See, e.g., *Malloy v. Hogan*, 378 U.S. 1 (1964).

124. See *Gideon v. Wainwright*, 372 U.S. 335 (1963).

125. See, e.g., *Coates v. City of Cincinnati*, 402 U.S. 611 (1971); *Cohen v. California*, 403 U.S. 15 (1971); *Brandenburg v. Ohio*, 395 U.S. 444 (1969); *Cantwell v. Connecticut*, 310 U.S. 296 (1940); *De Jonge v. Oregon*, 299 U.S. 353 (1937).

126. In philosophical language, the ontology of our constitutional criminal law is composed of this dual essence (*i.e.*, an embodiment of popular will and of the Bill of Rights). These two realities comprise its intrinsic nature. Thus, any test designed to reflect this essence must recognize and reflect that intrinsic tension in its full scope. Any test

Both embody critical democratic imperatives: respect for majority will and respect for those fundamental restrictions on state sovereign power that are also the glory of the American political experience. This intrinsic structural tension requires that determinations of guilt or innocence by the application of criminal statutes to individual cases respect and incorporate these twin imperatives. If, instead, the judicially-crafted implementing tests are centered exclusively on protection of the rights, interests, and values exemplified in the Bill of Rights, those tests slight or ignore the important governmental interests and values served by the legislature's exercise of the police power. Such one-sided tests do not authorize analysis and decisionmaking that is proportionate to the nature and scope of the twin legal realities at issue. They therefore fail what might be called a test for tests: they squeeze formulation and analysis so that an intrinsic two-sided reality is reduced to a one-sided reality.

Yet, the opposite premise is equally true. If judicially-crafted tests are centered exclusively on formulation and analysis designed to protect the police power interests and values served by criminal statutes, these tests fail to respect and implement the core rights, interests, and values protected by the Bill of Rights. These tests also fail the test for tests: they too squeeze formulation and analysis so that the intrinsic two-sided reality is reduced to a one-sided reality. In both instances, one-sided tests are reductive and distort analysis and decisionmaking.

The glaring *Smith* defect, therefore, is the Court's embracing of its one-sided test which focuses exclusively on the important governmental interests served by typical criminal statutes (unless the statute facially burdens religious practice). This one-sided test turns the constitutional framework of our democratic criminal law on its head by mandating a police power framework for assessing free exercise claims. In the Court's view, the importance of the governmental interest expressed in a routine criminal prohibition is automatically sufficient to overcome the fundamental interest expressed in the free exercise clause. In fact, the governmental interest in any typical criminal statute is sufficient, without any further analysis, to exclude the possibility of a violation of the free exercise clause. The Court's holding is perfectly clear: "[I]f prohibiting the exercise of religion . . . is not the object . . . but merely the incidental effect of a generally applicable and otherwise valid provision, the First Amendment has not been offended."¹²⁷ Thus, the *Smith* holding, because it focuses almost exclusively on the statute at stake, causes the text of

that overlooks part of that essence is reductive. The scientific norm of isomorphism also captures the identical requirement: the form or structure of a scientific analysis must be proportionate to the form or structure of that which is investigated.

127. Employment Div. v. Smith, 110 S. Ct. 1595, 1600 (1990).

the free exercise clause to fade away from our attention and lose its vitality.¹²⁸ The *Smith* holding also obliterates the policy underlying this Bill of Rights test: that the free exercise of religion, a preferred constitutional activity, is explicitly singled out by the text of the first amendment itself for special protection. The *Smith* test stands "in the place of the Constitution."¹²⁹

The *Smith* holding transforms the traditional framework. The free exercise clause is plainly subordinated to the mere existence of an otherwise valid criminal statute. This first amendment guarantee is hostage to any ordinary criminal statute.

D. A Comparison of the *Smith* Rule With Other First Amendment Tests

A comparison of the *Smith* Court's prohibition of free exercise analysis contrasts strikingly with the multitude of tests that require first amendment freedom of expression analysis when an issue arises in the application of a criminal statute. The talismanic power attributed by the Court to such an ordinary criminal statute is starkly absolutist. It is not repeated in other first amendment areas. Nor is the requirement repeated that the state must intentionally intrude on protected first amendment interests before a claim can be considered.

Before comparing the *Smith* approach with the abundance of freedom of expression tests, it is necessary to appreciate the role of tests in respecting and implementing the ideals of the first amendment. The majestic words of the first amendment are not self-applying.¹³⁰ The plethora of tests in first amendment cases are rationalized as necessary to effectuate the deliberately imprecise language of the amendment itself.¹³¹ These tests provide standards of review that enable lawyers and judges to enforce the guarantee of rights by applying the open-ended language of the text to diverse cases in a consistent and coherent manner. In principle, at least, the values, interests, and rights embodied in the first amendment are vindicated, and those who are similarly situated are treated equally. Judicially-crafted tests transform and illuminate the first amendment language from a statement of abstract ideals and goals, an inspiring democratic manifesto, to a repertoire of standards for lawyerly

128. See *id.* at 1608 (O'Connor, J., concurring).

129. See Nagel, *The Formulaic Constitution*, 84 MICH. L. REV. 165, 187 (1985).

130. "[E]ven so stout a libertarian as Alexander Meiklejohn has chided those who insist that the words 'abridging the freedom of speech or of the press' are 'plain words, easily understood.'" Mendelson, *On the Meaning of the First Amendment: Absolutes in the Balance*, 50 CALIF. L. REV. 821, 821 (1962).

131. For a scholarly critique of the Court's utilization of proliferating and layered tests in constitutional analysis, see Nagel, *supra* note 129.

argument and judicial decisionmaking. In jurisprudential language, the first amendment articulates broad principles that have been implemented and defined by tests in a wide range of cases over an extensive period of time.

1. Criminal Statutes and Content-Based Regulation of Free Expression: Track One Categorical Analysis.—Governmental restrictions of expression fall within two general areas, content-based restrictions and content-neutral restrictions, and thus, there are two corresponding modes of analysis. We begin with content-based restrictions and the corresponding mode of analysis. In this track one realm, criminal statutes have no talismanic power to exclude analysis of the first amendment rights, values, and interests at stake in drawing the boundaries for the few traditional exceptions to the “principle that government may not prescribe the form or content of individual expression.”¹³² These exceptions include the prohibition of advocating violent overthrow of the government, of “fighting words,” and of obscenity. The governing tests for these areas are expressed in categorical or *per se* rules that carefully define the behavior at issue: the *Brandenburg* clear and present danger test,¹³³ the *Chaplinsky* fighting words test,¹³⁴ and the *Miller* obscenity test.¹³⁵ These categorical tests shelter first amendment values, interests, and rights that arise in the application of criminal statutes while narrowly defining behavior that is beyond first amendment protection or behavior that may be criminally prohibited as a valid exercise of police power.

132. L. TRIBE, *supra* note 7, at 832. Actually, there are a good many forms of expression that are outside the scope of the protection of the first amendment. These include verbal and other agreements spelling out conspiracy, solicitation of a crime, restraint of trade, perjury, larceny by false pretense and by trick, and intentional infliction of emotional distress. “Federal securities regulation, mail fraud statutes, and common-law actions for deceit and misrepresentation are only some examples of our understanding that the right to communicate information of public interest is not ‘unconditional.’” *Curtis Publishing Co. v. Butts*, 388 U.S. 130, 150 (1967) (plurality opinion). Nevertheless, when one considers these forms of “expression” in light of the purposes of first amendment protection of free speech (that a free government, the quest for truth in the marketplace of ideas, and the needs of self-realization, all require the right of free speech), it is easy to understand that these crimes and torts are outside of the scope of protection of the first amendment guarantee.

The incontrovertible fact that certain forms of expression are not within the guarantee of free expression addresses the question of the scope of the guarantee, but not the question of the nature of the obligation for those forms of expression found to be within the guarantee. The failure to distinguish between “scope and obligation” is “sometimes astonishing.” Frantz, *supra* note 121, at 1436. “The premise is that the first amendment cannot be ‘absolute’ in the sense of unlimited in scope. But, the conclusion is that it cannot be ‘absolute’ in the sense of unconditionally obligatory within its proper scope, whatever that may be.” *Id.* Frantz aptly calls this conclusion a *non sequitur*.

133. *Brandenburg v. Ohio*, 395 U.S. 444 (1969).

134. *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942).

135. *Miller v. California*, 413 U.S. 15 (1973).

a. The clear and present danger test

In first amendment challenges to criminal prosecutions for advocating the overthrow of the government, the *Brandenburg* clear and present danger test permits limitation of the content of free and robust expression only if the expression subject to the criminal statute and the actual expression at issue is: (a) “directed to inciting or producing imminent lawless action” and (b) “likely to incite or produce such action.”¹³⁶ The *Brandenburg* test affirms the right of free expression even for advocacy directed at overthrowing our political, economic, and social institutions, except in the extreme and narrow situation in which both of its strict standards are met. Only then does the important governmental interest against imminent violent overthrow embodied in the criminal statute prevail over first amendment interests.

The narrowness of the *Brandenburg* test implicitly defines all other advocacy directed at transforming our institutions as protected free expression. Such hard core free expression trumps any police power argument that advocacy of hated and subversive ideas that threaten cherished premises of our political and civil society should be barred. Once advocacy is classified as protected under the central first amendment meaning of “no law . . . abridging the freedom of speech, or of the press,” there is no balancing with a governmental police power interest. The interest is simply subordinated to the preferred interest inherent in the command of the first amendment.

b. The fighting words test

Second, criminal prohibitions of fighting words are limited by the strict first amendment test that such statutorily prohibited epithets must “by their very utterance inflict injury or tend to incite an immediate breach of the peace” and the actual words used must “have a direct tendency to cause acts of violence by the person to whom, individually, the remark is addressed.”¹³⁷ Again, this fighting words test affirms the right of robust, even insulting, free expression, unless the strict standards for fighting words are met. In this narrow instance, the important police power interest in public peace and security prevails. As in the *Brandenburg* test, the intentional narrowness of the fighting words test implicitly defines all other insulting speech that cannot be squeezed into its restricted terrain as protected, hard core free expression. Once so classified, the first amendment interest in such speech simply defeats outright the police power

136. *Brandenburg*, 395 U.S. at 447.

137. *Houston v. Hill*, 482 U.S. 451, 461-62 (quoting *Gooding v. Wilson*, 405 U.S. 518, 525 (1972)).

claim that such speech should be barred or, at a minimum, be subject to balancing analysis.

c. *The obscenity test*

Third, the *Miller* obscenity test, which is usually applied to assess state criminal prohibitions of obscenity, focuses on the police power interest: “whether the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest” and “whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law.”¹³⁸ The third element of the test captures the first amendment sheltering of free expression by also requiring that “the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.”¹³⁹ The first and second elements of the test reflect the “legitimate [state] interest in prohibiting dissemination or exhibition of obscene material.”¹⁴⁰ In contrast to these two elements for obscenity, the third part of the test captures the first amendment interest in free expression by negatively defining obscenity as “lack[ing] serious literary, artistic, political, or scientific value.”¹⁴¹

The trifurcated *Miller* test, therefore, incorporates the free expression interest as an element of the test. While the first two elements reflect the police power interest in prohibiting obscenity, there can be no finding of obscenity unless the third element is also demonstrated (*i.e.*, unless the prosecution proves beyond a reasonable doubt that the allegedly obscene material “lacks serious literary, artistic, political, or scientific value”).¹⁴² If the third element is not proved, the prosecution fails and the police power interest is subordinated to the free expression interest. Though frequently criticized¹⁴³ and clearly no panacea for this perplexing area, the *Miller* test reflects the Court’s intent to restrict the definition of obscenity so that free expression is also protected.

138. *Miller v. California*, 413 U.S. 15, 24 (1973) (citations omitted).

139. *Id.*

140. *Id.* at 18. For a specification of these interests, see T. EMERSON, TOWARD A GENERAL THEORY OF THE FIRST AMENDMENT 89 (1966) [hereinafter T. EMERSON, GENERAL THEORY].

141. *Miller*, 413 U.S. at 24.

142. *Id.*

143. See, e.g., *Paris Adult Theatre I v. Slaton*, 413 U.S. 49, 84 (1973) (Brennan, J., dissenting) (“[N]one of the available formulas . . . can reduce the vagueness to a tolerable level. . . . [W]e are manifestly unable to describe [obscenity] in advance except by reference to concepts so elusive that they fail to distinguish clearly between protected and unprotected speech.”).

d. Trumping police power interests versus Smith absolutism

The important police power interests and values exemplified in the criminal statutes in prohibiting imminent violent overthrow, fighting words, and obscenity, although respected and promoted, do not automatically trump first amendment rights, interests, and values. In addition, these embodied police power interests do not foreclose formulation and consideration of relevant first amendment claims. These tests also do not require a discriminatory intention manifested on the face of the statute as a *sine qua non* for considering a first amendment claim. To the contrary, these three tests authorize free expression trumping of police power interests. They define and classify which expression falls within and without the protection of freedom of speech or of the press. They thereby specify the reach of hard core protected free expression. These tests, although characterized as categorical or *per se* rules rather than as balancing tests, incorporate what has been called "definitional balancing."¹⁴⁴ Each definition (clear and present danger, obscenity, or fighting words) and each application exemplifies the twin imperatives inherent in our system of constitutional criminal law.¹⁴⁵ Thus, these tests are two-sided categorical rules.

144. Tribe aptly describes definitional balancing:

Although only the case-by-case approach of track two takes the form of an explicit evaluation of the importance of the governmental interests said to justify each challenged regulation, similar judgments underlie the categorical definitions on track one. Any exclusion of a class of activities from first amendment safeguards represents an implicit conclusion that the government interests in regulating those activities are such as to justify whatever limitation is thereby placed on the free expression of ideas. Thus, determinations of the reach of first amendment protections on either track presuppose some form of "balancing" whether or not they appear to do so. The question is whether the "balance" should be struck for all cases in the process of framing particular categorical definitions, or whether the "balance" should be calibrated anew on a case-by-case basis.

L. TRIBE, *supra* note 7, at 792-93.

145. The power of the track one definitional approach to protected free expression can be summarized as:

The definer - the judge who undertakes to assign some distinct meaning to the constitutional proposition - has now drawn a line. It may be a wavering and uncertain line at many points. He may come up against cases which compel him to conclude that he has drawn it in the wrong place and that it should be moved. . . . [B]orderline cases can still arise which could arguably be placed on either side. Yet, despite all these difficulties, something new emerges from the mere fact that a line has been drawn. There are now cases that are not borderline: cases that are well within the line, as well as others well outside it. The definer has therefore placed limits . . . on his own future freedom of decision. There are cases in which (unless he is willing to change the rule or evade it with sophistries) he *must* say that freedom of speech has been unconstitutionally abridged, as well as others in

The *Smith* holding is also a “categorical” or *per se* rule. It too defines the behavior it classifies as within or without the protection of the first amendment. Yet, in striking contrast to the traditional two-sided categorical rules that shelter free expression interests, the unilateral *Smith* test rejects first amendment analysis and mandates that any police power interest automatically trumps any free exercise interest whenever an ordinary criminal statute is applied. Thus, the one-sided categorical rule postulated in *Smith* is an absolutist police power rule. There is virtually no way the free exercise interest can prevail, and its rationale is the very antithesis of the driving rationale of track one analysis, an area in which free expression absolutists essentially prevail.¹⁴⁶

2. Criminal Statutes and Content-Neutral Regulation of Free Expression: Track Two Balancing Analysis.

a. Track two governmental restriction

The second major way the government regulates free expression is through content-neutral restrictions that trigger the corresponding mode of track two analysis.¹⁴⁷ In this track two realm, the government does not seek directly to regulate the content of free expression by defining, as in track one, the scope of protected free expression. Instead, the government furthers a clearly valid and specific police power interest in public peace, safety, health, or order through criminal and civil statutes,

which he must say it has not. The definer, in other words, must ultimately give the constitutional proposition a certain amount of content which he regards as being obligatory on the court. Consequently, in cases falling clearly within the defined area, the definer is largely relieved of responsibility for results in particular instances which he may find personally distasteful.

Frantz, *supra* note 121, at 1435. In effect, Frantz argues that the categorical approach serves the core legal model purposes in a constitutional framework of *stare decisis*, adjudication by an impartial judge, reasonable predictability and certainty, and equal justice for those who are similarly situated. Frantz contrasts the definitional approach with *ad hoc* balancing by a judge:

For him, there can be no clearly protected area — all areas are subject to invasion whenever “competing interests” are sufficiently compelling. Furthermore, his initial assumption — without which he could never justify balancing — is that the constitutional proposition contained in the first amendment is incapable of being assigned any meaning which would not be too broad (or too narrow) for consistent application. Therefore, it must have been intended to be subject to unstated exceptions, which the court must make. . . . The *ad hoc* balancer’s constitution is empty until the court decides what to put into it. It does not speak until the court speaks for it. It is inherently incapable of saying anything to the judge.

Id.

146. See L. TRIBE, *supra* note 7, at 792.

147. *Id.*

but limits the right to free expression in the application of such statutes, sometimes inadvertently and sometimes intentionally. In Justice O'Connor's words, "Our free speech cases similarly recognize that neutral regulations that affect free speech values are subject to a balancing, rather than categorical, approach."¹⁴⁸ Such clearly valid regulations include generally applicable, facially neutral criminal statutes and ordinances prohibiting trespass, breach of the peace, littering, conspiracy, and picketing near courthouses. Applications of statutes that raise free expression issues occur in cases involving reasonable time, place, and manner restrictions,¹⁴⁹ speech,¹⁵⁰ symbolic speech,¹⁵¹ overbroad and vague statutes,¹⁵² and associational rights.¹⁵³

148. Employment Div. v. Smith, 110 S. Ct. 1595, 1612 (1990) (O'Connor, J., concurring).

149. See, e.g., Grayned v. City of Rockford, 408 U.S. 104 (1972) (ordinance prohibiting picketing around elementary school buildings except for peaceful school labor disputes found unconstitutional because school grounds cannot be declared off limits); Adderley v. Florida, 385 U.S. 39 (1966) (a conviction pursuant to a facially neutral trespass statute for a demonstration on that part of the jail grounds reserved for jail uses against the explicit objection of the sheriff responsible for the jail upheld as applied as a valid neutral regulation of the location of the protest); Marsh v. Alabama, 326 U.S. 501 (1946) (facially neutral trespass statute held not enforceable against the distribution of religious literature on streets of company town).

150. See, e.g., United States v. Spock, 416 F.2d 165 (1st Cir. 1969) (reversing convictions pursuant to a facially neutral conspiracy prohibition as applied to protected advocacy opposing the Vietnam War).

151. See, e.g., United States v. Eichman, 110 S. Ct. 2404 (1990) (reversal of convictions upheld pursuant to a congressional statute prohibiting the destruction of an American flag as applied to persons who burned a flag as part of a protest against governmental policy); Texas v. Johnson, 491 U.S. 397 (1989) (application of a facially neutral Texas criminal statute prohibiting the desecration of venerated objects violated the first amendment as applied to Johnson who burned an American flag during a political demonstration); Clark v. Community for Creative Non-Violence, 468 U.S. 288, 293 (1984) (application of a facially neutral federal park regulation prohibiting unauthorized camping in federal parks upheld as applied to demonstrators sleeping in Lafayette Park; the Court assumed, *arguendo*, that the overnight sleeping was "expressive conduct protected to some extent by the First Amendment").

152. The state interest and the free expression interest are balanced in these cases. If the state interest is compelling, the free expression interest is subordinated and less precision is mandated. See, e.g., Parker v. Levy, 417 U.S. 733 (1974) (Captain Levy's free speech rights in urging enlisted men to disobey orders to go to Vietnam were subordinated to a compelling military wartime interest in order and discipline, and his conviction for "conduct unbecoming an officer and a gentleman" was upheld). If the free expression interest is weighed against a modest or dubious state interest, more precision is required. See, e.g., Houston v. Hill, 482 U.S. 457 (1987) (ordinance prohibiting interrupting policeman in the execution of his duties invalidated as overbroad; however, "a properly tailored statute" could prohibit obstructing an investigation, creating a traffic hazard, and other behaviors typically prohibited by disorderly conduct statutes).

153. See, e.g., Talley v. California, 362 U.S. 60 (1960) (conviction based on ordinance prohibiting the distribution of handbills which did not contain the name and address of the author, printer, and sponsor reversed as a restraint on freedom of association).

b. Track two, case-by-case balancing detailed

The central task of track two analysis is choosing between the state interest and the first amendment interest.¹⁵⁴ In Tribe's words, "[I]t is impossible to escape the task of weighing the competing considerations."¹⁵⁵ Indeed, it is impossible for the conscientious judge to blink at her oath to uphold the twin imperatives of "the Constitution and laws of the United States."¹⁵⁶ The choice "between competing interests" is "struck on a case-by-case basis."¹⁵⁷ There is case-by-case balancing by weighing the particularized governmental interest embodied in regulatory criminal and civil statutes on their face and as applied against the specific free expression interest raised by the facts and the individual's claim. This case-by-case balancing can be complex. Yet, although tests vary depending on the particularized interests in conflict, there is no abject deference to any police power interest incorporated in the ordinary criminal statute and no automatic subordination of free expression rights,

154. Aleinikoff specifies two distinct forms of balancing: "Sometimes the Court talks about one interest *outweighing* another. Under this view, the Court places the interests on a set of scales and rules the way the scales tip. . . . Constitutional standards requiring 'compelling' or 'important' state interests also exemplify this form of the balancing metaphor." Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943, 946 (1987) (emphasis in original). This first form of balancing is distinguished from "'striking a balance' between or among competing interests. The image is one of balanced scales with constitutional doctrine calibrated according to the relative weights of the interests. One interest does not override another; each survives and is given its due." *Id.* In addition, Aleinikoff distinguishes balancing from methods of adjudication that look at a variety of factors in reaching a decision:

These would include some of the familiar multi-pronged tests and "totality of the circumstances" approaches. These standards ask questions about how one ought to characterize particular events. Was the confession voluntary or involuntary? . . . In answering . . . one starts with some conception of what constitutes voluntariness and involuntariness and then asks whether the particular situation shares more of the voluntary elements or the involuntary elements. . . . The reasoning is thus primarily analogical. Balancing represents a different kind of thinking. The focus is directly on the interests or factors themselves. Each interest seeks recognition on its own and forces a head-to-head comparison with competing interests.

Id. at 945.

Justice Black opposed balancing in principle and argued an absolutist approach that the literal first amendment language that Congress shall make "no law" abridging freedom of speech literally "means no law" without "any ifs, buts, or whereases." Cahn, *Justice Black and First Amendment "Absolutes": A Public Interview*, 37 N.Y.U. L. REV. 549, 553, 559 (1962). See also *Barenblatt v. United States*, 360 U.S. 109, 140-41 (1959) (Black J., dissenting); T. EMERSON, *GENERAL THEORY*, *supra* note 140, at 56-58.

155. L. TRIBE, *supra* note 7, at 792.

156. 28 U.S.C. § 453 (1988) (oaths of justices and judges).

157. L. TRIBE, *supra* note 7, at 792.

interests, and values. As in track one categorical analysis, the mere application of any typical criminal statute does not create talismanic power to exclude consideration of free expression when the government, directly or indirectly, intrudes on free expression.

c. The weaknesses of case-by-case balancing

Measured by the requirements of the positivist-dominated legal model for formulating, analyzing, and decisionmaking, the balancing approach to resolving competing interests, while not invalid, ranks low in the universe of modes of legal reasoning. The core legal model requirement that judges and jurors should be impartial decisionmakers, guided in their analysis by rules and principles that will govern the resolution of like cases in the future, is not well served by the ad hoc nature and elasticity of balancing. In a typical form of balancing, the judge in each case must initially identify the nature and weight of the particular police power interest at stake and then weigh that interest and determine whether the government used the least drastic means for attaining the specific police power interest.¹⁵⁸ This identification and weighing must be performed again to determine the precise first amendment interest at issue. The competing interests must then be weighed against each other with a determination that one interest outweighs the other in each case.¹⁵⁹

The immediate analytical questions are: How is such identifying, weighing, comparing, and determining to be done? What are the criteria for these tasks? Are these tasks for judges or is the court asked "to assess after each incident a myriad of facts, to guess at the risks created by expressive conduct, and to assign specific value to the hard-to-measure worth of particular instances of free expression"?¹⁶⁰ The balancing approach itself does not supply adequate answers to these questions and forces judges to supply their own answers by looking to their predilections in ways which are quite different from their application of the element-centered rules characteristic of track one analysis.

Neither does the aggregation of prior ad hoc balancing decisions, what has been called a "quagmire of ad hoc judgment,"¹⁶¹ aid such analysis and decisionmaking in new cases by providing a repertoire of explanations and justifications that has the authority of stare decisis and that therefore provides a framework for guidance. Nor does such an aggregation embody even a modest approximation of certainty and predictability, the absence of which could lead to chilling the exercise of

158. See *id.* at 977-86.

159. *Id.*

160. *Id.* at 793.

161. *Id.* at 794.

free expression because a person is never sure whether her free expression right will prevail over the governmental interest until an appellate court decides. In short, balancing is "a slippery slope . . . a matter of degree [and] first amendment protections become especially reliant on the sympathetic administration of the law."¹⁶²

d. The justification of case-by-case balancing

First, in Tribe's words, in the track two realm, "the 'balancers' are right in concluding that it is impossible to escape the task of weighing the competing considerations,"¹⁶³ and ad hoc balancing is better than "a constitutional system in which . . . governmental behavior would automatically be upheld, however devastating its consequences for free-

162. *Id.* Nimmer, in arguing against the use of what he calls "ad hoc balancing," has three objections.

First, ad hoc balancing by hypothesis means that there is no rule to be applied, but only interests to be weighed. In advance of a final adjudication by the highest court a given speaker has no standard by which he can measure whether his interest in speaking will be held of greater or lesser weight than the competing interest which opposes his speech. . . . [I]f there is no rule at all then there is no certainty at all. The absence of certainty in the law is . . . particularly pernicious where speech is concerned because it tends to deter all but the most courageous (not necessarily the most rational) from entering the market place of ideas.

Nimmer, *The Right to Speak from Times to Time: First Amendment Theory Applied to Libel and Misapplied to Privacy*, 56 CALIF. L. REV. 935, 939 (1968). Nimmer's second objection to case-by-case balancing is that it leads to favoring "the side which opposes freedom of speech." *Id.* at 940. The reason is that free speech issues will arise from "those who espouse the most unpopular ideas, those against whom feelings run the highest . . . and only they are likely to be prosecuted." *Id.* When judges "engage in the 'delicate and difficult task' of weighing competing interests," they are likely to be influenced by "strong popular feelings." *Id.* His third reason is that "the ingrained judicial deference to the legislative branch can and has in ad hoc balancing tended toward judicial abdication in the weighing process." *Id.* at 941. For a detailed critique of ad hoc balancing, see *Konigsberg v. State Bar of Cal.*, 366 U.S. 36, 60-61 (1961) (Black, J., dissenting); T. EMERSON, GENERAL THEORY, *supra* note 140, at 53-56; Frantz, *supra* note 121, at 1424.

Nimmer's critique of ad hoc balancing is a prelude to his argument that the balancing process should occur "on the definitional rather than the litigation or ad hoc level." Nimmer, *supra*, at 942. The level or place of balancing is changed so that the question is not "which litigant deserves to prevail in a particular case," but the broader question of "defining which forms of speech are to be regarded as 'speech' within the meaning of the first amendment." *Id.* Nimmer illustrates his argument with the Court's landmark decision in *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964), which determined that knowingly and recklessly false speech about public officials is not the type of speech protected by the first amendment. See Nimmer, *supra*, at 942-55. For a critique of Nimmer's argument that definitional balancing creates a rule which can be applied in the future without additional balancing, see Aleinikoff, *supra* note 154, at 979.

163. L. TRIBE, *supra* note 7, at 792.

dom of expression.”¹⁶⁴ The essence of balancing requires that both the police power interest and the first amendment interest be considered in formulating and analyzing competing interests, even though that analysis can lead to subordinating one interest to the other in decisionmaking. Thus, balancing promotes fidelity to the judge’s oath to uphold both the Constitution and the laws.¹⁶⁵

Second, balancing has a potential over time to create definitional rules.¹⁶⁶ The *Brandenburg* definitional rule, for example, evolved in 1969 in common-law fashion after decades of federal and state sedition cases controlled by extremely elastic balancing approaches that strongly favored the police power interest and usually resulted in convictions and affir-

164. *Id.* at 978.

165. According to W. Mendelson:

It is largely because of the absence of defining standards, I suggest, that the Court has resorted openly to balancing in free speech cases. We have had too many opinions that hide the inevitable weighing process by pretending that decisions spring full-blown from the Constitution — a document written generations ago by men who had not the slightest conception of the world in which we live. . . . Open balancing compels a judge to take full responsibility for his decisions, and promises a particularized, rational account of how he arrives at them — more particularized and more rational at least than the familiar parade of hollowed abstractions, elastic absolutes, and selective history. Moreover, this approach should make it more difficult for judges to rest on their predispositions without ever subjecting them to the test of reason. It should also make their accounts more rationally auditable.

Mendelson, *supra* note 130, at 825-26. Since this article was published in 1962, the Court has refined and increased the number of “defining standards,” including the *Miller* obscenity test, the *Brandenburg* version of the “clear and present danger” test, and the *New York Times* “defamation” test for public officials. Thus, the characterization of the alternative to balancing as comprised of “hallowed abstractions, elastic absolutes, and selective history” now has, in my judgment, somewhat less validity. Compared to ad hoc balancing, the definitional approach is superior in protecting free expression for the reasons Nimmer, Mendelson, and others detail. Balancing is also superior to “hallowed abstractions, elastic absolutes, and selective history.” For example, in *Di Santo v. Pennsylvania*, 273 U.S. 34 (1927), the Court decided that a state regulation was unconstitutional as a “direct burden” on interstate commerce without specifying “the decisive considerations that marked the burden in question as ‘direct’ rather than ‘indirect.’” See Mendelson, *The First Amendment and the Judicial Process: A Reply to Mr. Frantz*, 17 VAND. L. REV. 479, 482 (1964). For a historical analysis of the development of balancing and a worthwhile critique, see Aleinikoff, *supra* note 154.

166. Balancing suggested a particularistic, case-by-case, common law approach that accommodated gradual change and rejected absolutes. The outcome of a case would turn on a careful analysis of the particular interests at stake. Today, the plaintiff might win because of the unjustified burden imposed by a governmental regulation; tomorrow, the government could demonstrate an adequate public interest to sustain its legislation. Balancing could keep everyone in the game. It thus provided flexibility without sacrificing legitimacy.

Aleinikoff, *supra* note 154, at 961.

mations of such convictions on appeal.¹⁶⁷ The emergence of the narrower, element-dominated *Brandenburg* rule is closely correlated with the fading of these sedition prosecutions in favor of tolerance of a much more robust and exuberant free expression, even of hated and disturbing ideas.¹⁶⁸

Third, ad hoc judging with elasticity in formulating, analyzing, and deciding is for good as well as for ill. It allows a judge to seek and to capture more of the human complexity that is at stake without being fenced in by element-dominated rules. Positivism, in constitutional law as elsewhere, has limitations as well as advantages, and broad policy analysis can produce impressive examples of the legal imagination at work as well as personal and status quo myopia. Lastly, in many areas, balancing may simply be the best form of analysis that is possible at a particular time in deciding conflicts that cannot be anticipated precisely. This includes the array of conflicts that occur when the government, while aiming at the noncommunicative aspect of behavior, nevertheless significantly impacts free expression.¹⁶⁹ In these cases, balancing enables the judge to decide the controversy before her in as reasoned a manner as is possible at the time even though there is no guiding body of developed case law. The judge, who ordinarily must decide the instant case, does not have the luxury of the social philosopher who may defer resolution of issues until reflection has deepened her insight or until the available mode of reasoning has been refined in an evolving body of philosophical analysis.

e. *Track two analysis versus Smith absolutism*

The track two free expression cases referred to in this discussion unfold the precise police power interest posed and the precise first

167. See *Scales v. United States*, 367 U.S. 203 (1961); *Dennis v. United States*, 341 U.S. 494 (1951); *Whitney v. California*, 274 U.S. 357 (1927); *Gitlow v. New York*, 268 U.S. 652 (1925); *Abrams v. United States*, 250 U.S. 616 (1919); *Schenck v. United States*, 249 U.S. 47 (1919).

168. See, e.g., *NAACP v. Claiborne Hardware Co.*, 458 U.S. 886 (1982); *Communist Party v. Whitcomb*, 414 U.S. 441 (1974).

169. "Even first amendment absolutists accept something like a balancing or reasonableness standard in most time, place and manner contexts." C.E. BAKER, *supra* note 57, at 315 n.1. Justice Black, a leading advocate of the absolutist approach, saw no alternative to balancing in this context. See, e.g., *Konigsberg v. State Bar of Cal.*, 366 U.S. 36, 68-69 (1961) (Black, J., dissenting); *Barenblatt v. United States*, 360 U.S. 109, 141-42 (1959) (Black, J., dissenting); *Martin v. City of Struthers*, 319 U.S. 141, 143-44 (1943); Kalven, *The Concept of the Public Forum: Cox v. Louisiana*, 1965 SUP. CT. REV. 1, 28. Professor Thomas Emerson, the preeminent scholarly proponent of first amendment absolutism, recognized that in allocating the use of physical facilities for the exercise of the right of free expression, "[t]he governing principle can only be a fair accommodation of opposing interests [that is] a kind of balancing test." T. EMERSON, *THE SYSTEM OF FREEDOM OF EXPRESSION* 359 (1970).

amendment interest at stake by utilizing tests that focus analysis on these twin interests with some form of balancing to aid in resolving which interest shall prevail. In contrast, the one-sided *Smith* test mandates that any police power interest, no matter how modest, codified in an ordinary criminal statute *ipso facto*, trumps any free exercise interest, no matter how substantial, without any balancing at all. These free expression cases provide facial analysis of content neutrality and “as applied” analysis of whether neutrality has been demonstrated. *Smith* authorizes only facial analysis and emphatically rejects any case-by-case appraisal so that the entire terrain of intended and unintended harm to free exercise in application is not judicially cognizable.

The track two cases referred to in this discussion also have no *sine qua non* requirement that the government manifest its intent to intrude on protected free expression on the face of the statute before the first amendment claim can be adjudicated. In contrast, *Smith* imposes exactly this requirement as a *sine qua non*: that the government make plain on the face of the statute its intent to burden religious practice before the free exercise claim can be considered. In the wide array of track two cases, the variety of implementing tests applied sometimes results in protection of the governmental interest embodied in the criminal statute and other times results in protection of the first amendment interest.¹⁷⁰ The *Smith* test will almost always result in rejection of the free exercise interest in favor of the governmental interest. Lastly, in the track two cases previously discussed, there is a demonstration of respect and seriousness for the free expression claims, even when these claims are ultimately rejected after judicial scrutiny. This is in sharp contrast with *Smith*’s angry language in repudiating almost all free exercise claims that arise in a criminal context.¹⁷¹

170. See, e.g., *United States v. Eichman*, 110 S. Ct. 2404 (1990) (flag burning on the steps of the United States Capitol was expressive conduct protected by the first amendment and could not support convictions under the Flag Protection Act of 1989); *Texas v. Johnson*, 491 U.S. 397, 400 (1989) (reversing a conviction for burning an American flag as part of a protest upheld as a vindication of the defendant’s interest in freedom from content-based regulation of core free expression conduct); *Clark v. Community for Creative Non-Violence*, 468 U.S. 288, 294 (1984) (prohibition of “expressive sleeping” as part of a protest in violation of a ban on camping upheld as a reasonable “time, place or manner restriction or as a regulation of symbolic conduct”); *Spence v. Washington*, 418 U.S. 405, 411 (1974) (conviction for displaying the American flag upside down from the window of an apartment reversed as a “prosecution for the expression of an idea through activity”); *United States v. O’Brien*, 391 U.S. 367, 382 (1968) (conviction for knowingly burning a draft card during the Vietnam War upheld as a vindication of “[t]he governmental interest . . . [in] preventing a harm to the smooth and efficient functioning of the Selective Service System”).

171. See *supra* notes 52-55 and accompanying text.

E. Balancing and Free Exercise

1. *The Reynolds Draining of the Free Exercise Clause.*—In the nineteenth century landmark case of *Reynolds v. United States*,¹⁷² the Supreme Court articulated a belief/action distinction, rather than a balancing test, in assessing the validity of free exercise claims.¹⁷³ Reynolds, a Mormon, violated a federal criminal prohibition against polygamy in what was then the Utah territory of the United States. The *Reynolds* test distinguished freedom of religious belief and expression of religious opinion, which received first amendment protection, from freedom of action based on religious motivation, which could be regulated by a secular, otherwise valid statute.¹⁷⁴ Thus, a person's action in marrying

172. 98 U.S. 145 (1879).

173. *Id.* at 166.

174. *Id.* at 166-67. Justice Scalia, in writing for the *Smith* majority, reflected the *Reynolds* holding and rationale. His initial explication of the history of free exercise doctrine emphasized that the clause "means, first and foremost, the right to believe and profess whatever religious doctrine one desires. Thus, the First Amendment obviously excludes all 'governmental regulation of religious beliefs as such.'" *Employment Div. v. Smith*, 110 S. Ct. 1595, 1599 (1990). Justice Scalia also elaborated a second protected category: "performance of (or abstention from) physical acts: assembling with others for worship service, participating in sacramental use of bread and wine, proselytizing, abstaining from certain foods or certain modes of transportation." *Id.* Scalia, however, rejected a third category by excluding from free exercise protection all other behavior that has a "religious motivation." He wrote:

Respondents . . . seek to carry the meaning of "prohibiting the free exercise [of religion]" one large step further. They contend that their religious motivation for using peyote places them beyond the reach of a criminal law that is not specifically directed at their religious practice, and that is concededly constitutional as applied to those who use the drug for other reasons.

Id. Scalia cited *Reynolds* to refute the claim that "an individual's religious beliefs excuse him from compliance with an otherwise valid law prohibiting conduct that the State is free to regulate." *Id.* at 1600. Such laws, in the *Reynolds* language, "are made for the government of actions, and while they cannot interfere with mere religious beliefs and opinions, they may with practices." *Id.*

Justice O'Connor replied with a more expansive vision of what action is within the protection of "the express textual mandate" of the free exercise clause:

A person who is barred from engaging in religiously motivated conduct is barred from freely exercising his religion. . . . It is difficult to deny that a law that prohibits religiously motivated conduct, even if the law is generally applicable, does not at least implicate First Amendment concerns. . . . The First Amendment . . . does not distinguish between laws that are generally applicable and laws that target particular religious practices. . . . Our free exercise cases have all concerned generally applicable laws that had the effect of significantly burdening a religious practice.

Id. at 1608 (O'Connor, J., concurring).

The majority's extremely narrow definition of the scope of religious activity that is cognizable within the free exercise of religion manifests an extremely narrow and aggressively

an additional wife, though religiously praiseworthy, could violate a valid criminal prohibition of polygamy.

The *Reynolds* belief/action distinction was driven by a deep-seated, ethnocentric repugnance for polygamy as an evil to be suppressed,¹⁷⁵ by a sweeping view of legislative police power embodied in generally applicable statutes,¹⁷⁶ and by an either/or mode of reasoning that excludes and expresses alarm about any exceptions to general laws.¹⁷⁷ In Professor Lupu's words:

secularist conception of religion. It emphasizes beliefs and includes religiously motivated conduct, but only in the sense of rituals and ceremonies such as congregate worship, use of sacramental wine, proselytizing, and abstaining from certain foods. What is excluded by this secularist definition is all other religiously laden conduct beyond belief and rituals.

Wisconsin v. Yoder, 406 U.S. 205 (1972), specifically captures this broader idea of religion. In *Yoder*, the refusal of Amish parents to send their children to a public high school was recognized by the Court as religious based action within the purview of free exercise protection. Chief Justice Burger, speaking for the *Yoder* Court, stated:

[O]ur decisions have rejected the idea that religiously grounded conduct is always outside the protection of the Free Exercise Clause. . . . [T]o agree that religiously grounded conduct must often be subject to the broad police power of the State is not to deny that there are areas of conduct protected by the Free Exercise Clause of the First Amendment and thus beyond the power of the State to control, even under regulations of general applicability.

Id. at 219-20. The *Smith* Court's narrow view of the scope of conduct protected by the free exercise clause is also plainly inconsistent with the Court's validation of the religiously motivated refusal of a Jehovah's Witness to work on military tanks. *See Thomas v. Review Bd.*, 450 U.S. 707 (1981).

It is hardly surprising that a broad array of religious groups have expressed dismay at the *Smith* Court's rejection of the compelling interest test which proposes a narrow secularist conception of the scope of religious practice to be recognized by federal and state courts for free exercise analysis and decisionmaking. *See supra* note 8. The Court has given a new meaning to the ancient imperative "[t]o render unto Caeser the things that are Caeser's." *Matthew* 22:15-21. Caeser has a new magnitude of power. The state has authoritatively defined religious exercise as essentially private (composed only of beliefs and rituals) and has explicitly defined other religious-laden conduct as beyond first amendment protection.

175. *Reynolds*, 98 U.S. at 164 ("Polygamy has always been odious among the Northern and Western Nations of Europe and . . . was almost exclusively a feature of the life of Asiatic and African people.").

176. *Id.* (man "has no natural right in opposition to his social duties") (quoting a letter written by Thomas Jefferson). Jefferson's views on the scope of free exercise appear to be an exception to the general belief. "[T]he freedom of religion was almost universally understood (with Jefferson being the preeminent exception) to include conduct as well as belief." McConnell, *supra* note 8, at 1451-55.

177. *Reynolds v. United States*, 98 U.S. 145, 167 (1979) ("To permit [exceptions] would . . . in effect . . . permit every citizen to become a law unto himself. Government could exist only in name under such circumstances."). *Reynolds* illustrates the power of isolated legal formulation and analysis in narrowing the values and interests at stake in the existence of a religious people. In Tribe's words, "[f]ew decisions better illustrate how

Reynolds drained the free exercise clause of its primary constitutional function. Those forms of religiously motivated action that are also speech are already protected by the constitutional clauses concerned with freedom of expression. . . . *Reynolds'* belief-action distinction thus reduced the free exercise clause to a primarily rhetorical commitment to protecting religious liberty.¹⁷⁸

2. *The Fading of the Reynolds Belief/Action Dichotomy.*—Thus, for almost eighty years, the *Reynolds* belief/action dichotomy provided a thorny hurdle for most free exercise claims.¹⁷⁹ The reign of the *Reynolds* belief/action test came to an end, however, in *Braunfeld v. Brown*¹⁸⁰ in 1961 and in *Sherbert v. Verner*¹⁸¹ in 1963. In *Braunfeld*, orthodox Jewish merchants raised a free exercise claim against laws that forced them to close on Sunday after fidelity to their orthodox religious beliefs required them to close on Saturday.¹⁸² In a four week period, therefore, their stores would be open twenty days while their non-Sabbatarian competitors would be open twenty-four days, resulting in an economic burden on the Jewish merchants because of their adherence to their religious beliefs. Though deciding against the free exercise claim, the *Braunfeld* Court specifically rejected the contention that indirect governmental intrusions on the free exercise of religion do not violate the first amendment.¹⁸³ The Court stressed that intrusion into religious practice could flow from both governmental “purpose or effect”:

amorphous goals may serve to mask religious persecution.” L. TRIBE, *supra* note 7, at 1271. The history of the Mormons in the nineteenth century is filled with religious inspired persecution including violence, imprisonment, expulsion, and confiscation of property. *Id.* The *Reynolds* formulation and analysis is blind to this context. For Mormons, *Reynolds* is a footprint in a long trail of religious persecution. The *Smith* majority also formulated and analyzed in narrow fashion, as if its analysis and decision could be isolated from the historical trail of religious persecution of Native Americans referred to by Justice Blackmun in his dissent. *Smith*, 110 S. Ct. at 1622 n.10. For a detailing of this historical trail, see Barsh, *The Illusion of Religious Freedom for Indigenous Americans*, 65 OR. L. REV. 363, 369-74 (1986).

178. Lupu, *Where Rights Begin: The Problem of Burdens on the Free Exercise of Religion*, 102 HARV. L. REV. 933, 938 (1989).

179. The contradiction between the belief/action dichotomy and the text of the first amendment guarantee of the free exercise of religion is portrayed by Tribe. He states, “It is somewhat peculiar that the distinction between belief and action would persist in the free exercise context, for the guarantee refers explicitly to the *exercise* of religion, and thus seems to extend by its own terms beyond thought and talk.” L. TRIBE, *supra* note 7, at 1183 n.33 (emphasis in original).

180. 366 U.S. 599 (1961).

181. 374 U.S. 398 (1963).

182. *Braunfeld*, 366 U.S. at 600-01.

183. *Id.* at 607 (“[T]o hold unassailable all legislation regulating conduct which

If the purpose or effect of a law is to impede the observance of one or all religions or is to discriminate invidiously between religions, that law is constitutionally invalid even though the burden may be characterized as being only indirect. But if the State regulates conduct by enacting a general law within its power, the purpose and effect of which is to advance the State's secular goals, the statute is valid despite its indirect burden on religious observance unless the State may accomplish its purpose by means which do not impose such a burden.¹⁸⁴

In applying its least-draastic-means balancing test, the *Braunfeld* Court decided that anticipated difficulties for the police in enforcing the claimed exemption for Saturday Sabbatarians and a strong state interest in providing a uniform day of rest for all workers justified the denial of the exemption and the application of the closing laws to these merchants.¹⁸⁵ What is doctrinally significant in *Braunfeld*, however, is first, the implied rejection of the *Reynolds* belief/action dichotomy and the justification of its sweeping police power regulation of religious-based action by a generally applicable statute. Second, *Braunfeld* is also doctrinally significant in affirmatively legitimizing religious-based action as capable in principle of raising a free exercise claim that must be subjected to a balancing test. Even a valid "general law within its [police] power"¹⁸⁶ is scrutinized for direct and indirect burdens on religious practice that may raise such a claim.

3. *The Rise of Free Exercise: The Emergence of the Sherbert Compelling Interest Test.*—The *Sherbert* Court confirmed the overthrow of the *Reynolds* belief/action dichotomy and its justification of sweeping police power regulation of religious-based action. In place of the belief/action distinction, *Sherbert* clearly established the compelling interest balancing test.¹⁸⁷

imposes solely an indirect burden on the observance of religion would be a gross oversimplification.").

184. *Id.*

185. *Id.* at 608-09.

186. *Braunfeld v. Brown*, 366 U.S. 599, 607 (1961).

187. In his opinion for the Court, Justice Brennan emphasized:

In *Speiser v. Randall* we emphasized that conditions upon public benefits cannot be sustained if they so operate, whatever their purpose, as to inhibit or deter the exercise of First Amendment freedoms. . . . "To deny an exemption to claimants who engage in certain forms of speech is in effect to penalize them for such speech." Likewise, to condition the availability of benefits upon this appellant's willingness to violate a cardinal principle of her religious faith effectively penalizes the free exercise of her constitutional liberties.

Sherbert v. Verner, 374 U.S. 398, 405-06 (1963) (quoting *Speiser v. Randall*, 357 U.S. 513, 518 (1958)).

Sherbert involved a Seventh-Day Adventist in South Carolina who was fired because she would not work on Saturday, the Sabbath Day of her religion. The Court held that South Carolina could not constitutionally apply its facially neutral statutory test to force a worker to abandon his religious convictions respecting the day of rest.¹⁸⁸ The Court's holding followed its application of the compelling interest test: only a "compelling state interest enforced in the eligibility provisions of the South Carolina statute justifies the substantial infringement of appellant's First Amendment right."¹⁸⁹ The Court affirmed that even if it found a compelling interest, the state must still "demonstrate that no alternative forms of regulation would combat such abuses without infringing First Amendment rights."¹⁹⁰ The Court reiterated that "[i]f the purpose or effect of a law is to impede the observance of one or all religions . . . that law is constitutionally invalid even though the burden may be characterized as being only indirect."¹⁹¹

Lupu, a free exercise commentator, describes the doctrinal shift inherent in *Braunfeld* and *Sherbert* as "creat[ing] sweeping potential for the free exercise clause to become a source of judicial power to protect religious liberty against insensitivity as well as direct government hostility."¹⁹² *Sherbert* especially inaugurates the modern era of free exercise

188. *Id.* at 410.

189. *Id.* at 406.

190. *Id.* at 407. Michael W. McConnell, a historian of the free exercise clause, described the free exercise exemptions doctrine established in *Sherbert* in simple terms:

If the plaintiff can show that a law or governmental practice inhibits the exercise of his religious beliefs, the burden shifts to the government to demonstrate that the law or practice is necessary to the accomplishment of some important (or "compelling") secular objective and that it is the least restrictive means of achieving that objective. If the plaintiff meets his burden and the government does not, the plaintiff is entitled to exemption from the law or practice at issue. McConnell, *supra* note 8, at 1416-17.

191. *Sherbert v. Verner*, 374 U.S. 398, 404 (1963) (quoting *Braunfeld v. Brown*, 366 U.S. 599, 607 (1961)).

192. Lupu, *supra* note 178, at 942. Michael W. McConnell agrees: "The *Sherbert* decision . . . created the potential for challenges by religious groups and individual believers to a wide range of laws that conflict with the tenets of their faiths." McConnell, *supra* note 8, at 1412. The *Smith* Court appeared to react sharply against the potential of the *Sherbert* compelling interest test and rationale outside of the unemployment compensation area. In a scholarly decoding of free exercise history, McConnell concludes that "the historical evidence . . . does, on balance, support *Sherbert's* interpretation of the free exercise clause." *Id.* at 1415. In focusing on "exemptions from generally applicable laws," McConnell reaches three conclusions from his historical research:

(1) that exemptions were seen as a constitutionally permissible means for protecting religious freedom, (2) that constitutionally compelled exemptions were within the contemplation of the framers and ratifiers as a possible interpretation of the free exercise clause, and (3) that exemptions were consonant with the

jurisprudence.¹⁹³

4. *The Sherbert Progeny in Unemployment Compensation.*—What has become known as the *Sherbert* compelling interest test also controlled decisionmaking in *Thomas*, *Hobbie*, and *Frazee*. In all four unemployment compensation cases, the Court, in its formulation and analysis of the twin interests to be balanced, demonstrated respect for both the police power interest and the workers' claims. The Court particularized the government's administrative interests in uniform and efficient enforcement of the unemployment benefits statutes, and it described the free exercise claims of the minority religion members as both sincere and substantial. Indeed, there is no question raised in these cases as to the sincerity or substantiality of the claims that religious practice was burdened.¹⁹⁴

In applying the requirement that the burdening of religious practice be substantial before free exercise rights are violated, the Court's analysis acknowledged what was at stake for the workers. In *Thomas*, a Jehovah's Witness and steel mill worker quit his job when he was reassigned to work on the production of military tanks. Writing for the Court majority, Chief Justice Burger vividly captured the substantial interest at stake for this Jehovah's Witness as posing "a choice between fidelity to religious belief or cessation of work."¹⁹⁵ Citing *Sherbert*, the Chief Justice applied the compelling interest test and reiterated what the Court had stressed elsewhere: a "regulation neutral on its face may, in its application, nonetheless offend the constitutional requirement for governmental neutrality if it unduly burdens the free exercise of religion."¹⁹⁶

popular American understanding of the interrelation between the claims of a limited government and a sovereign God.

Id.

193. The absence of a detailed historical grounding of the *Sherbert* holding and rationale has been remedied by Michael McConnell's recent article. See McConnell, *supra* note 8.

194. *Frazee v. Illinois Dep't of Employment Sec.*, 489 U.S. 829, 833 (1989) ("We do not face problems about sincerity or about the religious nature of Frazee's convictions, however. The courts below did not question his sincerity, and the State concedes it. . . . [T]he Board of Review characterized Frazee's views as 'religious convictions.'"); *Hobbie v. Unemployment Appeals Comm'n*, 480 U.S. 136, 138 n.2 (1987) ("It is undisputed that appellant's conversion was bona fide and that her religious belief is sincerely held."); *Thomas v. Review Bd.*, 450 U.S. 707, 716 (1981) ("On this record, it is clear that Thomas terminated his employment for religious reasons."); *Sherbert v. Verner*, 374 U.S. 398, 399 n.1 (1963) ("No question has been raised in this case concerning the sincerity of appellant's religious beliefs. Nor is there any doubt that the prohibition against Saturday labor is a basic tenet of the Seventh-day Adventist creed.").

195. *Thomas*, 450 U.S. at 717 (working on producing military tanks "would be against all of the . . . religious principles that . . . I have come to learn").

196. *Id.* (quoting *Wisconsin v. Yoder*, 406 U.S. 205, 220 (1972)). From the perspective

In *Hobbie*, a recent convert to the Seventh-Day Adventist faith was fired when she could no longer work on Saturday and was refused unemployment benefits because of this "misconduct."¹⁹⁷ The Court reaffirmed *Sherbert* and *Thomas* and applied the compelling interest test.¹⁹⁸ The Court also emphasized that an "indirect" infringement "upon free exercise is nonetheless substantial."¹⁹⁹ The Court found "no meaningful distinction" among the issues posed in *Sherbert*, *Thomas*, and *Hobbie*.²⁰⁰

In *Frazee*, in 1989, a unanimous Court, applying *Sherbert*, *Thomas*, and *Hobbie*, invalidated a denial of unemployment benefits to a Christian who was not affiliated with any Church and who declined a work offer because the job required that he work on Sunday. The unemployment agency grounded its rejection of his claim on the fact that his belief was "personal" and not rooted in the "tenants or dogma" of a church or sect and that the Illinois courts had distinguished the Supreme Court's precedents in *Sherbert*, *Thomas*, and *Hobbie* on this basis.²⁰¹ There was no question of *Frazee*'s sincerity and, indeed, the state conceded it.²⁰² The Court summarily rejected the state's arguments.²⁰³

Sherbert, *Thomas*, *Hobbie*, and *Frazee* illustrate the power of the compelling interest balancing test to aid in promoting respect for and in scrutinizing the twin interests in conflict: the police power interest and the free exercise interest. They also illustrate the power of balancing analysis in a series of cases to create a narrow positivist rule that the statutory test of "without good cause" or "misconduct" may not be applied under the first amendment to deny benefits to a worker who quits or refuses work because her religion forbids work on the Sabbath Day of her faith, as well as a broader principle that a "choice between

of these Sabbatarians, there is, of course, nothing "neutral" about a requirement that mandates work on their Sabbath. The concept of neutrality is itself not neutral. It presupposes a perspective that defines and gives meaning to the word. For Sabbatarians, the claim is not simply individual. Its meaning arises from the duties they owe both to God and to Caesar.

197. *Hobbie v. Unemployment Appeals Comm'n*, 480 U.S. 136, 138 (1987).

198. *Id.* at 141. ("Both *Sherbert* and *Thomas* held that such infringements must be subjected to strict scrutiny and could be justified only by proof by the State of a compelling interest.").

199. *Id.* (quoting *Thomas v. Review Bd.*, 450 U.S. 707, 718 (1987)).

200. *Id.*

201. *Frazee v. Illinois Dep't of Employment Sec.*, 419 U.S. 829, 830 (1989).

202. *Id.* at 833. In *Frazee*, Justice White summarized *Sherbert*, *Thomas*, and *Hobbie* by stating, "each of the claimants had a sincere belief that religion required him or her to refrain from the work in question," and "[i]n each of these cases, the appellant was 'forced to choose between fidelity to religious belief and . . . employment.'" *Id.* at 832-33 (quoting *Hobbie v. Unemployment Appeals Comm'n*, 480 U.S. 136, 144 (1987)).

203. *Id.*

fidelity to religious belief or cessation of work”²⁰⁴ forces a worker to abandon her religious convictions. In contrast, the principle articulated by the *Smith* Court violates both the letter and spirit of these precedents. It is an example of what Justice Blackmun, in dissenting in *Smith*, calls a “distorted view of our precedents.”²⁰⁵ Notwithstanding the unmistakably dominant theme in these cases of preferring, even celebrating, the claimants’ free exercise interests rather than the state’s weak administrative interests, the *Smith* majority’s begrudging reduction is that “our decisions in the unemployment cases stand for the proposition that where the State has in place a system of individual exemptions, it may not refuse to extend that system to cases of ‘religious hardship’ without compelling reason.”²⁰⁶ Thus, the *Smith* Court viewed the free exercise interest through the distorting prism of its police power myopia.²⁰⁷

5. *The Apogee of the Sherbert Test in Yoder.*—Outside of the unemployment compensation cases, the promise of the *Sherbert* compelling interest balancing approach in the free exercise area was realized in *Wisconsin v. Yoder*.²⁰⁸ In this landmark case, the Court applied the *Sherbert* balancing approach to a facially neutral, generally applicable criminal statute that required all children to attend public or private school until the age of sixteen.²⁰⁹ Two sets of unrelated Amish parents refused to send their children, ages fourteen and fifteen, to public school after the eighth grade. The parents were convicted of violating the compulsory attendance law and a small fine was imposed. In assessing the nature of the general police power interest raised by these facts, the Court described the interest in promoting education as “at the very apex of the function of a State.”²¹⁰

Nevertheless, applying *Pierce v. Society of Sisters*,²¹¹ Chief Justice Burger found that “even this paramount responsibility was, in *Pierce*, made to yield to the right of parents to provide an equivalent education in a privately operated system.”²¹² Burger affirmed that “a State’s interest in universal education, however highly we rank it, is not totally free from a balancing process when it impinges on fundamental rights and interests, such as those specifically guaranteed by the Free Exercise Clause

204. *Thomas v. Review Bd.*, 450 U.S. 707, 717 (1981).

205. *Employment Div. v. Smith*, 110 S. Ct. 1595, 1616 (Blackmun, J., dissenting).

206. *Id.* at 1603.

207. See *supra* notes 97-99 and accompanying text.

208. 406 U.S. 205 (1972).

209. *Id.* at 207 n.2. From an Amish perspective this statute is, of course, not neutral.

210. *Id.* at 213.

211. 268 U.S. 510 (1925).

212. *Yoder*, 406 U.S. at 213.

of the First Amendment.”²¹³ The Court specifically rejected the idea of an absolute police power interest in “universal compulsory education” and stated that “it is by no means absolute to the exclusion or subordination of all other interests.”²¹⁴

Although the Court found that Wisconsin’s interest in enforcing a system of compulsory education would be compelling in the “generality of cases,” its analysis rejected, on a variety of grounds, the application of this core police power interest to the claim.²¹⁵ The Court found the state’s interest in compulsory school attendance to be “one thing . . . when its goal is the preparation of the child for life in modern society . . . but it is quite another if the goal of education be viewed as the preparation of the child for life in the separated agrarian community that is the keystone of the Amish faith.”²¹⁶

In effect, the Court found that the rationale for the state statute was served by the Amish parents’ withdrawal of their children from school so that they could prepare for the Amish life through day-to-day participation in their family, communal, and farm existence, which was pervaded with religious meaning. In assessing the Amish free exercise interest, the Court meticulously detailed the impact of secular high school education on the Amish faith and upheld the trial court’s finding that the Amish parents sincerely believed that a secular high school could “endanger their own salvation and that of their children.”²¹⁷ The Court decided that enforcing the state’s compulsory education statute to require these Amish children to attend a secular high school would “gravely endanger if not destroy the free exercise of respondents’ religious beliefs.”²¹⁸ In its reasoning, the Court explicitly rejected the state’s argument that because the statute “applies uniformly to all citizens of the State and does not, on its face, discriminate against religions or a particular religion,”²¹⁹ it was constitutionally sound. Citing *Sherbert* and *Walz v. Tax Commission*,²²⁰ the Court stated, “A regulation neutral on its face may, in its application, nonetheless offend the constitutional requirement for governmental neutrality if it unduly burdens the free exercise of religion.”²²¹

Yoder illustrates the strength of ad hoc balancing when the facts and the nature of the conflict make the definitional approach both

213. *Id.* at 214.

214. *Id.* at 215.

215. *Id.* at 221-22.

216. *Id.* at 222.

217. *Id.* at 209.

218. *Id.* at 219.

219. *Id.* at 220.

220. 397 U.S. 664 (1970).

221. *Wisconsin v. Yoder*, 406 U.S. 205, 220 (1972).

inappropriate and unlikely. The Court's assessment of the state interest does not stop with the "paramount" police power interest in public education or with the neutrality of the state criminal statute, nor does the Court masquerade analysis with the confounding cloak of conclusory rhetoric about the direct or indirect impact of a statute. Chief Justice Burger impressively particularizes and unfolds this state interest and its impact on the Amish faith-laden rural life. The analysis also reveals a sensitive understanding and appreciation of the religious values inherent in the Amish faith and way of life, including the threat to free exercise of their religious beliefs.²²² In short, the application of the *Sherbert* compelling interest balancing test captures the heart and soul of what was at stake for the Amish (*i.e.*, the survival of their religion) and balances this core interest against the particular application of the general state interest in compulsory education. The scope of the *Sherbert* compelling interest balancing test permits and even promotes this in-depth analysis and is an impressive example of the legal imagination at work.

The *Yoder* Court's analysis is immune to criticism by advocates of the positivist form of the legal model. Given the extraordinary facts posed by the collision of a facially neutral state statute with the faith-driven Amish way of life, the possibility of a similar case and the possibility of extracting a positivist element-centered rule is modest. Because the facts in *Yoder* are *sui generis* or nearly so, the positivist need for a precedent, inspired by the needs of stare decisis in a constitutional context to meet the expectations of reasonable certainty and predictability, seems inapplicable. *Yoder* also illustrates that it is surely "impossible to escape the task of weighing the competing considerations,"²²³ unless one favors a "constitutional system in which . . . governmental behavior would automatically be upheld, however devastating its consequences."²²⁴ Surely balancing is better than this alternative. Though *Yoder* yields no simple positivist rule for the future, it exudes respect for the Amish free exercise interest, a respect that is an exemplar for the Court as it confronts other free exercise conflicts.

The contrast of *Yoder* with the *Smith* majority opinion is startling. *Yoder* exemplifies case-by-case individualization, the judicial function applied; *Smith* forbids case-by-case individualization. *Yoder* applies the compelling interest balancing test; *Smith* repudiates the compelling interest test and any balancing test at all. *Yoder* exhibits pervasive respect for

222. Chief Justice Burger eloquently summarized the vision permeating Amish culture: "Amish society emphasizes . . . a life of 'goodness,' rather than a life of intellect; wisdom, rather than technical knowledge; community welfare, rather than competition; and separation from, rather than integration with, contemporary worldly society." *Id.* at 211.

223. L. TRIBE, *supra* note 7, at 792.

224. *Id.* at 978.

the Amish free exercise claim which is detailed page after page;²²⁵ *Smith* repudiates the Native American free exercise claim with passionate language, and there are no pages, no paragraphs, not even a sentence, that respectfully details Smith's and Black's religious interest in the annual sacramental rite of their Native American Church. In *Yoder*, the state interest in the criminal statute mandating compulsory school attendance, although at the apex of state function and a paramount responsibility, is balanced against and subordinated to the Amish parents' free exercise interest.²²⁶ In *Smith*, the state interest in the application of the Oregon statute to the ingestion of peyote in a sacred rite of the Native American Church was so modest that Oregon never sought to prosecute Smith and Black and did not evince any concrete interest in enforcing its drug laws against religious users of peyote.²²⁷ In sum, *Yoder* celebrates the values and interests symbolized by the free exercise clause, while the absolutist holding in *Smith* suppresses most free exercise analysis and celebrates as triumphant the state police power interest in any typical criminal statute.

6. *The Decline of the Sherbert Test.*—Although since 1972, the Court has rejected all of the claims for a free exercise exemption presented to it (except for the unemployment compensation cases), the analysis in these cases differs sharply from the *Smith* analysis. Three cases applied the *Sherbert* compelling interest balancing test,²²⁸ two others rejected formal compelling interest balancing when special institutions are involved,²²⁹ one denied, in a Native American land case, that any cognizable free exercise interest existed,²³⁰ and one denied a free exercise interest in refusing to provide a child's social security number.²³¹ Yet, unlike *Smith*, no absolutist rule is applied in these cases that prohibits free exercise formulation and analysis. There is no *a priori* rejection of the validity of the free exercise claim as in *Smith* and no bar to the legitimacy of considering and adjudicating the free exercise interest. The subor-

225. See *Yoder*, 406 U.S. at 207-13.

226. *Id.* at 213.

227. Employment Div. v. Smith, 110 S. Ct. 1595, 1617 (1990) (Blackmun, J., dissenting).

228. *Hernandez v. Commissioner*, 490 U.S. 680 (1989); *United States v. Lee*, 455 U.S. 252 (1982); *Gillette v. United States*, 401 U.S. 437 (1971).

229. Justice O'Connor distinguished *Goldman v. Weinberger*, 475 U.S. 503 (1986) and *O'Lone v. Estate of Shabazz*, 482 U.S. 342 (1987), because "they arose in the narrow, specialized contexts in which we have not traditionally required the government to justify a burden on religious conduct by articulating a compelling interest." *Smith*, 110 S. Ct. at 1612 (O'Connor, J., concurring).

230. *Lyng v. Northwest Indian Cemetery Protective Ass'n*, 485 U.S. 439 (1988).

231. *Bowen v. Roy*, 476 U.S. 693 (1986).

dination of the free exercise interest in favor of the governmental interest follows the Court's scrutiny of the free exercise claim.²³²

Thus, in *Gillette v. United States*,²³³ the Court upheld the governmental interest in military conscription against a religious objection to participation in the "unjust" Vietnam War.²³⁴ Applying the *Sherbert* compelling interest balancing test, the Court, although affirming that Gillette was "guided by fundamental principles of conscience and deeply held views about the purpose and obligation of human existence,"²³⁵ decided that the burdens resulting from the "impact of the conscription laws" were justified by the governmental interest at issue.²³⁶ The Court again applied the *Sherbert* compelling interest test in *United States v. Lee*²³⁷ and rejected the free exercise claim of an Amish employer who refused to collect and pay social security taxes for his employees.²³⁸ The Court, while explicitly acknowledging both Lee's sincerity and the substantial burden on religion imposed in this situation, nevertheless justified the burden as "essential to accomplish an overriding governmental interest" in the uniform application of the social security system.²³⁹ The Court feared that allowing a religious exception to the social security tax could lead to many "exceptions flowing from a wide variety of religious beliefs."²⁴⁰

In *Hernandez v. Commissioner*,²⁴¹ the Court again applied the *Sherbert* compelling interest balancing test.²⁴² In applying this test, the Court rejected the claim that payments by members of the Church of Scientology for auditing and training meetings were deductible charitable contributions or gifts to the church under Internal Revenue Code section 170(c). The Court reaffirmed the principle that "even a substantial burden would be justified by the 'broad public interest in maintaining a sound tax system,' free of 'myriad exceptions flowing from a wide variety of religious beliefs.'"²⁴³ Unlike *Lee*, the religious interest in *Hernandez* did

232. In *Lyng*, however, the Court concluded that there is no free exercise interest at stake in the land claim of Native American plaintiffs. *Lyng*, 485 U.S. at 452.

233. 401 U.S. 437 (1971).

234. *Id.* at 439.

235. *Id.*

236. *Id.* at 461.

237. 445 U.S. 252 (1982).

238. *Id.* at 261.

239. *Id.* at 257-58.

240. *Id.* at 260.

241. 490 U.S. 680 (1989).

242. *Id.* at 699 ("The free exercise inquiry asks whether government has placed a substantial burden on the observation of a central religious belief or practice and, if so, whether a compelling governmental interest justifies the burden.").

243. *Id.* at 699-700 (quoting *United States v. Lee*, 455 U.S. 252, 260 (1982)).

not flow from a specific doctrinal obligation not to pay taxes, but from an argument that church members were forced to pay an incrementally larger tax burden because their payments for auditing were not classified as deductible contributions by the Internal Revenue Service.²⁴⁴ Nevertheless, Justice Marshall, writing for the Court, detailed and analyzed at length the exact nature and role of the free exercise claim before subordinating it to the governmental interest.²⁴⁵

In *Goldman v. Weinberger*,²⁴⁶ the Court rejected the free exercise claim of an Orthodox Jew who wore a small yarmulke while on duty in violation of an Air Force regulation mandating uniform dress.²⁴⁷ The Government conceded that Goldman's claim was sincere,²⁴⁸ and the Court did not dispute the importance of Goldman's religious interest in wearing his yarmulke.²⁴⁹ Nonetheless, because "the military is, by necessity, a specialized society separate from civilian society," the Court rejected the application of the compelling interest test and decided that "great deference" should be shown to the "professional judgment of military authorities concerning the relative importance of a particular military interest."²⁵⁰ Although the Court did not detail the free exercise claim and rejected formal balancing in this case, it considered both interests before deciding that the free exercise claim should clearly be subordinated to the military claim.²⁵¹

In *Bowen v. Roy*,²⁵² the Court held, eight to one, that the free exercise clause did not give the claimants a right to refuse to furnish the social security number of their two-year-old daughter. The Court stressed that an individual does not have a free exercise right "to require the Government to conduct its own internal affairs in ways that comport with the religious beliefs of particular citizens."²⁵³ Yet, at least five Justices indicated that in a properly framed case, they would impose an exemption from the statutory mandate that applicants for benefits

244. *Id.* at 700.

245. *Id.* at 698-700.

246. 475 U.S. 503 (1986).

247. *Id.* at 504. The yarmulke was five and one-half inches in diameter and dark colored. That the wearing of a yarmulke, a constant silent prayer, could provoke the torrent of justifying rhetoric from the Court about the need for "great deference to the professional judgment of military authorities concerning the relative importance of a particular military interest" is patently absurd, at least to anyone who has served in the military. *See id.* at 507.

248. *See id.* at 525 (Blackmun, J., dissenting).

249. *Id.* at 510 (Stevens, J., concurring).

250. *Id.* at 506-07.

251. *Id.* at 507-10.

252. 476 U.S. 693 (1986).

253. *Id.* at 699.

furnish their social security numbers.²⁵⁴ In a three member plurality analysis, Chief Justice Burger suggested a reasonableness test for challenging a governmental benefit “neutral and uniform in its application” and distinguished *Sherbert* and *Thomas* because the statutory “good cause” standard present in the unemployment compensation statutes “created a mechanism for individualized exemptions,”²⁵⁵ unlike the statutory requirement for furnishing a social security number which does not allow for exceptions.²⁵⁶ The Chief Justice reiterated the *Sherbert* Court’s reasoning that “[a] governmental burden on religious liberty is not insulated from review simply because it is indirect.”²⁵⁷

In *O’Lone v. Estate of Shabazz*,²⁵⁸ the Court, in a five-four decision, upheld the application of a prison regulation that barred Islamic prisoners from attending a Friday afternoon religious service.²⁵⁹ The Court’s analysis affirmed that prisoners retain limited free exercise protections²⁶⁰ and detailed both the sincerity of the religious claim²⁶¹ and its substantiality.²⁶² In this special prison context, the Court applied a reasonableness test and stated that “when a prison regulation impinges on inmates’ constitutional rights, the regulation is valid if it is reasonably related to legitimate penological interests.”²⁶³

In *Lyng v. Northwest Indian Cemetery Protective Association*,²⁶⁴ the Court rejected the free exercise claim of a Native American who protested the building of a six mile segment of road to connect with two other completed portions of a road in the Chimney Rock section of the Six Rivers National Forest.²⁶⁵ A study commissioned by the United States Forest Service recommended that the road not be completed because “constructing a road along any of the available routes ‘would cause serious and irreparable damage to the sacred areas which are an integral and necessary part of the belief systems and lifeway of Northwest California Indian peoples.’”²⁶⁶ After detailing the free exercise claim and

254. *Id.* at 715, 731-33.

255. *Id.* at 708.

256. *Id.*

257. *Id.* at 706.

258. 482 U.S. 342 (1987).

259. *Id.* at 345.

260. *Id.* at 348.

261. *Id.* at 345 (“There is no question that respondents’ sincerely held religious beliefs compelled attendance at Jumu’ah.”).

262. *Id.* at 351 (“[w]hile we in no way minimize the central importance of Jumu’ah to respondents”).

263. *Id.* at 349 (quoting *Turner v. Safley*, 482 U.S. 78, 89 (1987)).

264. 485 U.S. 439 (1988).

265. See *id.* at 442.

266. *Id.*

finding that the claim was sincere and substantial,²⁶⁷ the Court majority distinguished *Sherbert* and specifically rejected the application of the compelling interest test.²⁶⁸

Applying an unjustifiably broad analogical reading of *Bowen v. Roy*, the *Lyng* Court concluded that the instant “building of a road . . . cannot meaningfully be distinguished from the [governmental] use of a Social Security number in *Roy*,”²⁶⁹ and stressed that in both cases the “affected individuals would not “be coerced by the Government’s action into violating their religious beliefs; nor would either governmental action penalize religious activity by denying any person an equal share of the rights, benefits, and privileges enjoyed by other citizens.”²⁷⁰ While first stressing that “this Court has repeatedly held that indirect coercion or penalties on the free exercise of religion, not just outright prohibitions, are subject to scrutiny under the First Amendment,”²⁷¹ the *Lyng* Court concluded, paradoxically, that these holdings do “not and cannot imply that incidental effects of government programs, which may make it more difficult to practice certain religions but which have no tendency to coerce individuals into acting contrary to their religious beliefs, require government to bring forward a compelling justification for its otherwise lawful actions.”²⁷²

Although the Court explicitly acknowledged that the proposed road segment could have devastating effects on traditional Indian religious practices, Justice O’Connor, writing for the Court majority, found that “the Constitution simply does not provide a principle that could justify upholding respondents’ legal claims.”²⁷³ This conclusion seems to be driven by the Court’s apprehension that recognizing the Native American claim in *Lyng* could lead to other lawsuits in which Native Americans might “seek to exclude all human activity but their own from sacred areas of the public lands . . . [and] could easily require *de facto* beneficial ownership of some rather spacious tracts of public property.”²⁷⁴ The Court noted that the government cannot operate if it is required to satisfy every citizen’s religious needs and desires, and therefore, indi-

267. *Id.* at 447 (“It is undisputed that the Indian respondents’ beliefs are sincere and that the Government’s proposed actions will have severe adverse effects on the practice of their religion.”).

268. *Id.* at 450-52.

269. *Id.* at 449. The author, however, believes that this comparison is tortured on its face.

270. *Id.*

271. *Id.* at 450.

272. *Id.*

273. *Id.* at 452.

274. *Id.* at 452-53.

viduals should look to legislatures and other institutions to reconcile the various competing religious demands on government.²⁷⁵ The *Lyng* Court explicitly rejected the “respondents’ proposed extension of *Sherbert* and its progeny”²⁷⁶ in favor of the *Roy* rationale because *Roy* “offers a sound reading of the Constitution.”²⁷⁷ The *Lyng* Court’s conclusion may have been driven by the recognition that “the Government has taken numerous steps to minimize the very impact that construction of the road will have on the Indians’ religious activities.”²⁷⁸

The *Lyng* decision is similar to *Smith* in its readiness to drain the free exercise clause of meaning when a claim of burdening free exercise conflicts with a governmental regulation, in its dispatch to the legislatures of those seeking religious exemptions from governmental regulations,²⁷⁹ and in its utilization of a direct/indirect test for recognizing free exercise claims. Nevertheless, the *Lyng* opinion differs from *Smith* in that the *Lyng* Court considered and even detailed at length the nature of the particular Native American claim of burdening, stressed both the sincerity and substantiality of the claim, and strongly urged governmental respect and accommodation of the religious needs of Native Americans and other citizens.²⁸⁰

7. *The Redraining of the Free Exercise Clause: The Revival of Reynolds Belief/Action.*—The *Smith* Court’s evisceration of the free exercise clause creates a virtually impenetrable thicket for future free exercise claims. *Smith* revives the *Reynolds* police power absolutism approach by affirming a sweeping principle that any police power interest specified in an otherwise valid criminal statute defeats any free exercise claim and even bars consideration of free exercise claims by the courts. Like *Reynolds*, the *Smith* principle reduces this intrinsic two-sided reality to a one-sided reality. Despite decades of experience to the contrary, the *Smith* decision expresses an embedded fear that allowing courts to consider free exercise claims will precipitate a rush of claims and risk “courting anarchy.”²⁸¹ *Smith* revives the aggressively secularist belief/action dichotomy of *Reynolds* and affirms that the free exercise clause protects religious beliefs and the performance of religious acts such as abstaining from foods and proselytizing, but not all other action based on “religious motivation”²⁸² except when statutes incorporate a system

275. *Id.* at 452.

276. *Id.*

277. *Id.*

278. *Id.* at 440.

279. *Id.* at 452.

280. *Id.* at 453-54.

281. Employment Div. v. Smith, 110 S. Ct. 1595, 1605 (1990).

282. *Id.* at 1599.

of individual exemptions or a hybrid free exercise/free expression or parental right claim is presented.

Like *Reynolds*, *Smith* resolves the “theoretical and existential tension” that is at the heart of our Constitution. The *Smith* decision embodies respect for the majority will and the Bill of Rights and favors a vision of an American people that calls upon us to subordinate ourselves to a majoritarian decisionmaking that is encoded into generally applicable laws and is hostile to any exemptions. Yet, what is sacrificed is the imperative that respect for the Bill of Rights calls for a vision of us as a polyglot American people willing to carve into police power statutes a system of individual exemptions that includes respect for the free exercise of religion as well as for the rest of the Bill of Rights.

Lastly, unlike all the other cases except *Reynolds*, there is not a single word of respect in the *Smith* majority opinion for the undisputed sincerity and central significance of the free exercise claim of Alfred Smith and Galen Black, two members of the old Native American Church.²⁸³ Quoting *Lyng*, the majority dismissed the claimants’ religious interest as an interest in “spiritual development.”²⁸⁴ Smith’s and Black’s participation in this annual and ancient religious rite, “the *sine qua non* of [their] faith,”²⁸⁵ was reduced to a quest for “spiritual development.”²⁸⁶ This characterization was not inadvertent. Like the *Reynolds* rhetoric, the *Smith* majority decision is replete with angry words rejecting the Native American claim followed by an emphatic repudiation of any judicial role in considering such a claim.²⁸⁷ The contrast with the respect

283. See *id.* at 1622 (Blackmun, J., dissenting).

284. *Id.* at 1603.

285. People v. Woody, 61 Cal. 2d 716, 725; 394 P.2d 813, 820 (1964). In *Woody*, the California Supreme Court characterized the application of a California criminal statute to a Native American religious ceremony as tearing out the “theological heart of Peyotism.” *Id.* at 722, 394 P.2d at 818.

286. Employment Div. v. Smith, 110 S. Ct. 1595, 1603 (1990).

287. *Id.* at 1606 n.5. One of the cardinal justifications of the *Smith* Court’s rejection of a judicial role in adjudicating such claims is detailed by the Court majority:

Nor is it possible to limit the impact of respondents’ proposal by requiring a “compelling state interest” only when the conduct prohibited is “central” to the individual’s religion. It is no more appropriate for judges to determine the “centrality” of religious beliefs before applying a “compelling interest” test in the free exercise field, than it would be for them to determine the “importance” of ideas before applying the “compelling interest” test in the free speech field. What principle of law or logic can be brought to bear to contradict a believer’s assertion that a particular act is “central” to his personal faith? Judging the centrality of different religious practices is akin to the unacceptable “business of evaluating the relative merits of differing religious claims.” As we reaffirmed only last Term, “[i]t is not within the judicial ken to question the centrality of particular beliefs or practices to a faith, or the validity of particular litigants’

for the claims of the Seventh-Day Adventists in *Sherbert* and *Hobbie*, the Jehovah's Witness in *Thomas*, the unaffiliated Christian in *Frazee*, the Amish parents in *Yoder*, and the Native American in *Roy*, is especially

interpretation of those creeds." Repeatedly and in many different contexts, we have warned that courts must not presume to determine the place of a particular belief in a religion or the plausibility of a religious claim.

If the "compelling interest" test is to be applied at all, then, it must be applied across the board, to all actions thought to be religiously commanded. *Id.* at 1604-05 (citations omitted).

In commenting on Justice O'Connor's reference in her *Smith* concurrence to the instant free exercise burdening as "constitutionally significant" and to Justice Blackmun's reference in his dissent to "the severe impact of a State's restrictions on the adherents of minority religion" the *Smith* majority replied:

In any case, dispensing with a "centrality" inquiry is utterly unworkable. It would require, for example, the same degree of "compelling state interest" to impede the practice of throwing rice at church weddings as to impede the practice of getting married in church. There is no way out of the difficulty that, if general laws are to be subjected to a "religious practice" exception, *both* the importance of the law at issue *and* the centrality of the practice at issue must reasonably be considered. . . . [I]nquiry into "severe impact" is no different from inquiry into centrality. He [Blackmun] has merely substituted for the question "How important is X to the religious adherent?" the question "How great will be the harm to the religious adherent if X is taken away?" There is no material difference.

Id. at 1605 n.4 (emphasis in original).

This abstract analysis by the Court majority omits the plain reason why neither O'Connor nor Blackmun discussed the issue of substantiality at any length. In O'Connor's words:

There is *no dispute* that Oregon's criminal prohibition of peyote places a severe burden on the ability of respondents to freely exercise their religion. . . . As we noted in *Smith I*, the Oregon Supreme Court concluded that the Native American Church is a recognized religion, that peyote is a sacrament of that church, and that respondent's beliefs were sincerely held.

Id. at 1613 (O'Connor, J., concurring) (emphasis added). In Blackmun's words, "Respondents believe, and their sincerity has *never* been at issue, that the peyote plant embodies their deity, and eating it is an act of worship and communion. Without peyote, they could not enact the essential ritual of their religion." *Id.* at 1622 (Blackmun, J., dissenting) (emphasis in original).

The point is not that the Court majority's abstract analysis is without merit. Clearly, questions about substantiality of a free exercise claim could pose real difficulty, and the Court's arguments, though manifestly overdrawn and too sweeping, are important. Rather, the point is that these conceptual difficulties did not arise in *Smith* or in the array of landmark free exercise cases that are summarized by the Court. In *Thomas v. Review Board*, 450 U.S. 707, 714 (1981), for example, the Court accepted the claimant's explication of the burden from his perspective. The deeper point, however, is that conceptual acrobatics should be carried out in light of the facts and issues presented in such a body of case law. To engage in superficially impressive flights of abstraction emphasizing that there is invariably "no way out of the difficulty" in the face of numerous decisions that reveal *no dispute* about this issue is to engage in analysis that lacks proportionality and authenticity.

vivid. In these cases, the diverse claimants remained at the center of the Court's formulation and analysis. In *Smith*, the Court's formulation and analysis cause Alfred Smith, Galen Black, and the central religious rite of their faith, to fade away to the periphery. In our legal culture, informed by belief in the dignity and worth of all people, such reasoning should be suspect. It is especially unfortunate that after the "many years of religious persecution and intolerance"²⁸⁸ of Native American religions by majoritarian institutions, the Court, which speaks for all of us, spoke so badly.

8. *The Smith Assault on Our Democratic Theory and Practice.*—The *Smith* repudiation of free exercise claims that arise within the criminal context is accompanied by an instruction that leaves accommodation to the political process in the legislature with explicit acknowledgment that this legislative "political process will place at a relative disadvantage those religious practices that are not widely engaged in."²⁸⁹ This disadvantage is described as "that unavoidable consequence of democratic government."²⁹⁰ Justice O'Connor aptly replied to this provocative re-interpretation of our theory and history of "democratic government" when she wrote that the "First Amendment was enacted precisely to protect the rights of those whose religious practices are not shared by the majority and may be viewed with hostility," and the "history of our free exercise doctrine amply demonstrates the harsh impact majoritarian rule has had on unpopular or emerging religious groups such as the Jehovah's Witnesses and the Amish."²⁹¹ She added the moving words of Justice Jackson in *West Virginia State Board of Education v. Barnette*:²⁹²

The very purpose of a Bill of Rights was to withdraw certain subjects from the vicissitudes of political controversy, to place them beyond the reach of majorities and officials and to establish

In addition, there is a diversion from the remaining conceptual problem, which deserves authentic analysis. For a discussion of centrality in free exercise claims, see L. TRIBE, *supra* note 7, at 1247-51.

288. *Smith*, 110 S. Ct. at 1622 (Blackmun, J., dissenting).

289. *Id.* at 1606.

290. *Id.* Pastor Richard John Neuhaus, a conservative Lutheran theologian, described this language as "one of the most callous statements in contemporary court language or in the recent history of court language." *Firing Line: An Extraordinary Supreme Court Decision* (PBS television broadcast, July 24, 1990) (transcript available through Southern Educational Communications Association, P.O. Box 5966, Columbia, S.C. 29250).

291. *Employment Div. v. Smith*, 110 S. Ct. 1595, 1613 (1990). As Michael McConnell points out, under one view of the threat that government poses to religious freedom, "[t]he evil includes not only active hostility, but also majoritarian presuppositions, ignorance and indifference." McConnell, *supra* note 8, at 1418.

292. 319 U.S. 624 (1943).

them as legal principles to be applied by the courts. One's right to life, liberty, and property, to free speech, a free press, freedom of worship and assembly, and other fundamental rights may not be submitted to vote; they depend on the outcome of no elections.²⁹³

Lastly, the *Smith* Court's repudiation of any need to scrutinize the free exercise claim is contradicted by the oath to uphold the first amendment command that the legislature "shall make no law . . . prohibiting the free exercise [of religion]."²⁹⁴ How can the Court decide whether a law prohibits the free exercise of religion without actually examining the specifics of the claim?

In contrast to other first amendment guarantees in both track one and track two realms and to most of the modest but important tradition of free exercise case law, the unmistakable nature and effect of the absolutist *Smith* holding and principle is to drain the meaning from the first amendment guarantee of free exercise of religion whenever the claim arises from the application of a typical criminal statute. It is an example of what Robert Cover calls judges "kill[ing] law."²⁹⁵ As he stated, "Confronting the luxuriant growth of a hundred legal traditions, [judges] assert that *this one* is law and destroy or try to destroy the rest."²⁹⁶ In

293. *Smith*, 110 S. Ct. at 1613 (O'Connor, J., concurring) (quoting *Barnette*, 319 U.S. at 638). "America's principal contribution to political theory is a conception of democracy according to which the protection of individual rights is a pre-condition, not a compromise, of that form of government." Dworkin, *The Reagan Revolution and the Supreme Court*, N.Y. Rev. of Books, July 18, 1991, at 23. "[T]he overriding virtue of and justification for vesting the Court with [the] awesome power [of judicial review] is to guard against governmental infringement of individual liberties secured by the Constitution." J. CHOPER, JUDICIAL REVIEW AND THE NATIONAL POLITICAL PROCESS: A FUNCTIONAL RECONSIDERATION OF THE ROLE OF THE SUPREME COURT 64 (1980). In Michael McConnell's words:

Locke's key assumption of legislative supremacy no longer holds under a written constitution with judicial review. The revolutionary American contribution to political theory was that the people themselves are sovereign and therefore possess inherent power to limit the power of the magistrate through a written constitution enforced by judges independent of the legislature and executive. . . . Once the courts are vested with the power to determine the proper boundary between individual conscience and the magistrate's authority . . . fuller protection for conscience becomes conceivable. . . . Once the people empowered the courts to enforce the boundary between individual rights and the magistrate's power, they entrusted the courts with a responsibility that prior to 1789 had been exercised only by the legislature.

McConnell, *supra* note 8, at 1444-45.

294. U.S. CONST. amend. I.

295. See Cover, *The Supreme Court 1982 Term, Foreward: Nomos and Narrative*, 97 HARV. L. REV. 4, 53 (1983).

296. *Id.* (emphasis in original).

Smith, the majority “kills” the potential for a more “luxuriant growth” from *Sherbert* and its progeny and asserts what is real: State police power is profoundly enlarged here to a new order of magnitude. Justice O’Connor’s characterization of this startling power as “talismanic” is surely apt, for an ordinary criminal statute embodying a police power interest has the magical power to forge a new constitutional criminal jurisprudence. Where there is such a statute, *Smith*, in effect, rewrites and transforms the two-hundred-year-old free exercise text:

Congress shall make no criminal law whose direct target is to prohibit the free exercise of religion.

or

Congress shall make no law prohibiting the free exercise of religion except where there is a typical, (*i.e.*, facially neutral, generally applicable, criminal statute).²⁹⁷

Smith transforms the state advocate armed with the weapon of an ordinary criminal statute into a legal Hercules with the power to hurl a free exercise claim from the legal arena.

IV. CRIMINAL STATUTES AND CRIME CONTROL

The *Smith* attribution of “talismanic” power to an ordinary criminal statute also finds no justification under our theory of crime control. There is no crime control justification in concentrating on across-the-board criminal statutes which define crimes without a concurrent concentration on interrelated across-the-board statutes, including those which define legal capacity as well as all other relevant defenses. There is also no crime control justification in concentrating on such statutes without also considering constitutional claims raised by the facts.

In our jurisprudence, crime control justifications are not autonomous. They exist within the frameworks established by our federal and state constitutions and by the requirements of just liability.²⁹⁸ To appreciate the Court’s error in *Smith*, it is important first to understand the

297. *Smith* creates yet a third possible reformulation: Congress shall make no law prohibiting the free exercise of religion provided that the free exercise claim is a “hybrid” (*i.e.*, raised “in conjunction with other constitutional protections, such as freedom of speech and of the press . . . or the right of parents”). See *Employment Div. v. Smith*, 110 S. Ct. 1595, 1601 (1990).

298. In philosophical terms, utilitarian justifications of punishment and general and individual deterrence are limited by retributive justifications of just punishment rooted in ideas of what the offender deserves. See, e.g., I. KANT, THE METAPHYSICAL ELEMENTS OF JUSTICE 99-102 (J. Ladd trans. 1965); W. LAFAVE & A. SCOTT, CRIMINAL LAW 62-190 (1986); P. LOW, R. BONNIE, & J. JEFFRIES, CRIMINAL LAW: CASES AND MATERIALS 4 (1986).

important role and purpose of criminal statutes in serving crime control interests. This exposition also leads to understanding the limitation of such statutes in serving these interests.

A. General Deterrence

From the perspective of general deterrence, the across-the-board prohibitions set forth in penal law statutes embody threats to those who may be inclined to commit specific offenses.²⁹⁹ If a person engages in behavior prohibited by the statute, that person is at risk for a finding of liability and for the specified punishment. In addition, the threats embodied in these statutes may exert a cumulative threat against violation of these norms.³⁰⁰ The threats posed by the penal law are therefore both particular and cumulative in deterring prohibited behavior. The fact that many offenders are not actually deterred does not invalidate these threats. The validity of these threats and the validity of general deterrence as a crime control theory does not require that everyone, or even that most people, conform their behavior to the penal law norms.³⁰¹

Institutional enforcement of the threats of penal law by police, prosecutors, courts, and prisons demonstrates that the threats are taken seriously in everyday life. Who will commit a crime with a police officer at his elbow? When the threats fail to deter, the theory of general deterrence provides a justification for punishment. The offender is punished as an example to others who may be tempted to violate these norms. The theory is that it is both just and useful to use the offender as an example. The theory is "just" because the offender was on notice that such punishment was threatened for failure to conform to the penal law norm and "useful" because others may be deterred.³⁰² The offender made a free choice to violate the norm and is responsible for that choice.³⁰³ In addition to the general deterrent effect, such punishment also promotes crime control under the related theory of individual deterrence. For example, the imprisonment of the offender inhibits the possibility of threats to the general public during the term of impris-

299. See, e.g., Andenaes, *The General Preventive Effects of Punishment*, 114 U.P.A. L. REV. 949 (1966); Bentham, *An Introduction to the Principles of Morals and Legislation*, in THE UTILITARIANS 170 (1961).

300. Andenaes, *supra* note 299, at 950-51.

301. *Id.* at 955.

302. See, e.g., O.W. HOLMES, THE COMMON LAW 41 (1963); Brandt, *Rule Utilitarianism*, in PHILOSOPHICAL PERSPECTIVE ON PUNISHMENT 93, 93-94 (G. Ezorsky ed. 1972).

303. See, e.g., O.W. HOLMES, *supra* note 302, at 41; Brandt, *supra* note 302, at 93-94.

onment, and the pain of imprisonment may also inspire law-abiding behavior in the future.³⁰⁴

In sophisticated articulations of general deterrence, these threats are also seen as reinforcing social norms inculcated by families, schools, churches, temples, and communities in the ordinary socialization process.³⁰⁵ The penal law threats strengthen this socialization process with an influential impact. This influence may be of greater significance than the direct threat of punishment to would-be offenders or actual punishment of those found liable for offenses.³⁰⁶ In our jurisprudence, these are the crime control purposes of across-the-board criminal prohibitions.³⁰⁷

Yet, there is no crime control justification in our tradition of criminal law jurisprudence for empowering a facially neutral, across-the-board charging statute to bar a constitutional claim. To articulate this notion is to expose its bankruptcy because all crime control efforts exist within a constitutional framework and are guided by it. More specifically, there is no crime control justification that flows from the principle that the mere existence of such a statute prevents a free exercise claim from being arguable and adjudicable. To the contrary, the values protected by these statutes are routinely balanced against the values protected by a range of constitutional guarantees and ordinary defenses.

B. Deterrence and Balancing

Indeed, general deterrence, when applied as a justification for individual punishment, compels case-by-case balancing of crime control values with a spectrum of competing values. These latter values are embodied in a system of individualized exemptions that require judges to determine that punishment is justified because the exemptions do not apply and liability is established. The deterrent justification of punishment is restricted to those who are adjudicated as offenders. It does not apply to those who, contrary to the initial charges and categorization, are

304. Delaney, *Towards a Human Rights Theory of Criminal Law: A Humanistic Perspective*, 6 HOFSTRA L. REV. 831, 878 (1978).

305. Andenaes, *supra* note 299, at 950, 956-57.

306. *Id.* at 978.

307. The utilitarian concept of crime control purposes is sometimes defined to include incapacitation, which can be viewed as part of individual deterrence, and rehabilitation, which is best viewed as a collateral objective of crime control and not itself an appropriate aim of punishment. See N. MORRIS & G. HAWKINS, LETTER TO THE PRESIDENT ON CRIME CONTROL 67-68 (1977); H. PACKER, THE LIMITS OF THE CRIMINAL SANCTION 50-51 (1968); A. VON HIRSCH, DOING JUSTICE: THE CHOICE OF PUNISHMENTS (1976). General criminal statutes play a role in applying these additional "purposes" in varying degrees, depending on what is authorized by the penal code and the actual behavior of the sentencing judge.

found not to be punishable or who demonstrate a justification or excuse for their behavior rooted in self-defense, defense of another, mistake of fact, entrapment, duress, necessity, or other defense, which when established, exculpates the defendant from liability for the charge flowing from the across-the-board criminal prohibition.

In addition, the deterrent justification of punishment does not apply equally to those whose capacity to choose is diminished (*i.e.*, the mentally impaired,³⁰⁸ the very young,³⁰⁹ the intoxicated,³¹⁰ and the provoked³¹¹). It discriminates between the degree of punishment believed necessary to send a message to first offenders and the punishment believed necessary to send a message to repeat offenders.³¹² Deterrence is simply subordinated to the vindication of competing constitutional interests when defendants who may have the required *mens rea* and *actus reus* are discharged from liability because their fourth, fifth, or sixth amendment rights were violated by governmental agents. We routinely subordinate the relevant crime control interest to the constitutional system of individualized exemptions. Crime control defers to the commanding interests embodied in the Bill of Rights.

C. Facial and Case-by-Case Analysis

Applying these basic jurisprudential principles, the crime control significance of the criminal statute injected by the *Smith* majority must be examined not only on its face, but also in light of its particularized application by the Court. Both "facial" and "as applied" analysis are necessary to determine the exact crime control interest at stake. As to facial analysis, the Oregon statute, which is aimed at users and prohibits the knowing or intentional possession of a controlled substance unless the substance was prescribed by a medical practitioner,³¹³ is on its face

308. Andenaes, *supra* note 299, at 958.

309. *Id.*

310. See, e.g., *People v. Hood*, 1 Cal. 3d 444, 462 P.2d 370, 82 Cal. Rptr. 618 (1969).

311. See, e.g., MODEL PENAL CODE § 4.01 (1962).

312. When utilized as a justification for sentencing, general deterrence provides a basis for individualizing the actual sentence imposed within the scope authorized by the across-the-board statute under which the defendant was found liable. Aggravating factors are balanced against ameliorating factors in determining the type of sentence that will serve to deter others. The rationale of general deterrence, as well as the rationale of just punishment, compels this crime control calibration. General deterrence will not be served by imposing an identical sentence on a first offender and an incorrigible offender. There is no sense in imposing an identical sentence on a person who stole \$1,500 and a person who stole \$15,000,000, although both may be liable for grand larceny. See Andenaes, *supra* note 299, at 960-70.

313. See OR. REV. STAT. §§ 475.992(4), 475.005(6), 475.992(4)(a), 475.035 (1987).

at the bottom of the specific hierarchy of crime control interests involved in controlling drugs. At the top are the statutes penalizing the sale or delivery of narcotics,³¹⁴ which are aimed at major drug suppliers and dealers. At an in-between level are the statutes prohibiting the possession of narcotics in sufficiently substantial amounts to create a statutory presumption of possession with the intent to sell or as classified in Oregon, attempted delivery.³¹⁵ Incidentally, this hierarchy of crime control interests manifestly reflects the political role of the criminal law in expressing the clear democratic will. Drug dealers, especially at higher echelons in the drug culture, bear a commonly accepted opprobrium in the community far greater than the street addict who buys small amounts to sustain his habit.

The *Smith* facts do not implicate a crime control interest in deterring the sale or delivery of narcotics or the possession of amounts sufficient to fall within the classification of attempted sale or delivery. The specific crime control interest raised by these remarkably special facts is limited to the use, and hence the possession, of apparently modest amounts of peyote by Smith and Black in the annual sacramental rite of their Native American Church. Thus, in Justice Blackmun's words, "[i]t is not the State's broad interest in fighting the critical 'war on drugs' that must be weighed against respondents' claim, but the State's narrow interest in refusing to make an exception for the religious, ceremonial use of peyote."³¹⁶ If the inquiry as to the state's interest is not appropriately narrowed, the result can be a loaded formulation of "the weighing process in the State's favor"³¹⁷ (e.g., the individual's interest in justice as balanced against the core state or public interest in crime control in the "war on drugs"). To avoid this loaded formulation, which leads to a foreordained conclusion in favor of the government, the interests weighed must be, in Roscoe Pound's words, "on the same plane . . . [or] we may decide the question in advance in our very way of putting it."³¹⁸

It is not possible to precisely identify the amount of peyote possessed by the two respondents in *Smith* for a reason that is revealing about the puny crime control interest inherent in these facts: "Oregon has

314. In Oregon, sales are included within a broader category of "delivery" of a controlled substance. *Id.* § 475.992(1).

315. In Oregon, these cases are captured within the prohibition of attempted delivery of a controlled substance. See, e.g., *Oregon v. Boyd*, 92 Or. App. 51, 756 P.2d 1276 (1988).

316. *Employment Div. v. Smith*, 110 S. Ct. 1595, 1617 (1990) (Blackmun, J., dissenting).

317. *Id.*

318. Pound, *A Survey of Social Interests*, 57 HARV. L. REV. 1, 2 (1943).

never sought to prosecute respondents, and does not claim that it has made significant enforcement efforts against other religious users of peyote."³¹⁹ Although such a charge could have been brought,³²⁰ this exercise of prosecutorial discretion reflects the extremely low crime control priority attached to the events in *Smith* by the local police and prosecutors who were charged with protection of crime control interests in the community. Possibly, the fact that there is no significant illicit traffic in peyote motivated this law enforcement response.³²¹ The drug menace confronting the country does not include peyote, especially not the sacramental use of peyote in an annual rite of a religion that is centuries old.³²² In Justice Blackmun's persuasive words: "The State cannot plausibly assert that unbending application of a criminal prohibition is essential to fulfill any compelling interest, if it does not, in fact, attempt to enforce that prohibition."³²³

D. Legislative Intent

Finally, the Oregon Supreme Court, in assessing the legislative intent underlying the Oregon unemployment compensation statute, clearly found that "the legality of ingesting peyote does not affect our analysis of the state's interest."³²⁴ The court made this finding of legislative intent and rejected the contrary determination of the Oregon Unemployment Compensation Board. The court stated that the relevant state interest "in denying unemployment benefits to a claimant discharged for religiously motivated misconduct must be found in the unemployment compensation statutes, not in the criminal statutes proscribing the use of peyote."³²⁵ The Oregon unemployment compensation statute, which provides monetary benefits to the unemployed, and the Oregon penal law, which seeks to control and punish crime, serve different purposes. The penal law purposes should not influence the unemployment statute.

Although he conceded that the United States Supreme Court is not bound by this judicial finding as to the intent of the Oregon legislature in enacting the unemployment statutes, Justice Brennan, in *Smith I*, commented, "we have never attributed to a state legislature a validating purpose that the State's highest court could find nowhere in the stat-

319. *Smith*, 110 S. Ct. at 1617 (Blackmun, J., dissenting).

320. *See id.* at 1597.

321. "In this case, the State actually has not evinced any concrete interest in enforcing its drug laws against religious users of peyote." *Id.* at 1617 (Blackmun, J., dissenting).

322. *See id.* at 1618-19.

323. *Id.* at 1617.

324. *Smith v. Employment Div.*, 301 Or. 209, 218-19, 721 P.2d 445, 450 (1986), *vacated*, 485 U.S. 660 (1988).

325. *Id.* at 219, 721 P.2d at 450.

ute.”³²⁶ Nevertheless, the *Smith* Court foisted its own view of the legislative intent underlying the Oregon unemployment compensation statute without commenting on the emphatic rejection of such a view by the Oregon Supreme Court, both initially and on remand. Such disrespect, even contempt, for the finding of the Oregon Supreme Court concerning traditionally state-controlled crime control interests is inconsistent with federalism. The majority opinion in *Smith* reflects an unprincipled activism, an aggressive reaching out to foist an exaggerated view of the uncharged and untried crime control interest at stake upon the Oregon Supreme Court, which strongly rejected the relevance of this particular interest in the *Smith* unemployment compensation context.³²⁷

Even worse, the *Smith* Court’s rejection of any balancing test forbids the calibration in case-by-case scrutiny of future cases of the exact crime control interest at stake which would then be balanced against the claimed burdening of religious practice. The general crime control interest in a criminal statute in any application is deemed to be automatically sufficient not only to override the claimed burdening, but also to bar consideration of the claim. The wide spectrum of crime control interests raised by various criminal statutes and their application to incredibly diverse facts are collapsed into one interest for this purpose. Any statute, any application, will do!

From the standpoint of crime control theory and practice, this is an extraordinary result because every day thousands of prosecutors throughout the country in hundreds of thousands of cases routinely balance the nature, scope, and magnitude of the crime control interest presented by actual cases and prospective investigations against scarce prosecutorial and judicial time and resources and then attach high, moderate, or low priority to a particular case or investigation on a spectrum of crime control significance.³²⁸

326. Employment Div. v. Smith, 485 U.S. 660, 677 (1988) (Brennan, J., dissenting).

327. Justice Blackmun, although reluctantly agreeing that the issue of the constitutionality of the Oregon statute criminalizing the use of peyote was “properly presented” in a technical sense in *Smith*, nevertheless was critical of the Court’s adjudication of this issue.

I have grave doubts, however, as to the wisdom or propriety of deciding the constitutionality of a criminal prohibition which the State has chosen not to enforce . . . and which the Oregon courts could, on remand, either invalidate on state constitutional grounds, or conclude that it remains irrelevant to Oregon’s interest in administering its unemployment benefits program. It is surprising, to say the least, that this Court which so often prides itself about principles of judicial restraint and reduction of federal control over matters of state law would stretch its jurisdiction to the limit in order to reach, in this abstract setting, the constitutionality of Oregon’s criminal prohibition of peyote use.

Employment Div. v. Smith, 110 S. Ct. 1595, 1616 n.2 (1990) (Blackmun, J., dissenting). The Court’s unprincipled and aggressive activism in *Smith* is further demonstrated by the fact that neither side argued for the abolition of the compelling interest test.

328. Supervising judges also do such weighing in assigning especially skilled judges

Sworn to uphold the laws, the prosecutor attaches priorities which reflect penal law categorical priorities (robberies receive higher priority than larcenies and grand larcenies receive higher priority than petit larcenies) and degrees of harm within such categories (the more shocking intentional murder rivets attention and receives the highest priority compared to the less shocking murder or manslaughter). On the basis of this routine balancing, prosecutors and investigators are assigned or not assigned to particular cases, supervisory and preparation time is prioritized, and grand jury presentations and trials are planned. This everyday crime control balancing has an ironic exception. There is no need to do such balancing when there is a free exercise claim. Any police power interest, no matter how modest, defeats any free exercise claim, no matter how intrusive.

Other than *ipse dixit* magic, why should the mere existence of this statutory prohibition of possession, which was never applied to Smith or Black and was twice emphatically rejected by the Oregon Supreme Court, have the power automatically to bar a claim of a burdening of religious practice? As a matter of foundational crime control principle, why should any crime control interest, no matter how petty, prevail against any free exercise interest, no matter how intrusive? The attribution in *Smith* of this magical statutory power in the face of the puny, uncharged, and untried crime control threat injected by the *Smith* Court contradicts our crime control jurisprudence as well as core premises of our theory of just liability and of our constitutional criminal law.

V. CONCLUSION

In our political culture, the United States Supreme Court has on occasion the “prophetic” function³²⁹ of summoning the nation to rally to its moral vision. It expounds “the best thinking” of what we “stand for as a people.”³³⁰ The Court should be “the voice of the spirit, reminding us of our better selves”³³¹ and calling the people to “provisional judgment—in the here and now.”³³²

to high priority crime control cases, and trial and appellate judges often must weigh whether to frame a case in light of a crime control model or a “due process” model. See Packer, *Two Modes of the Criminal Process*, 113 U. PA. L. REV. 1 (1964).

329. M. PERRY, THE CONSTITUTION, THE COURTS, AND HUMAN RIGHTS: AN INQUIRY INTO THE LEGITIMACY OF CONSTITUTIONAL POLICYMAKING BY THE JUDICIARY 98-99 (1982).

330. S. BARBER, ON WHAT THE CONSTITUTION MEANS 9 (1984).

331. A. COX, THE ROLE OF THE SUPREME COURT IN AMERICAN GOVERNMENT 117 (1976).

332. The secular “task of prophecy,” in Perry’s words, is to call “a people . . . to judgment—provisional judgment—in the here and now.” M. PERRY, *supra* note 329, at 98-99. Perry continues:

In the beginning . . . Americans have interpreted their history as having religious

By institutionalizing this prophetic function, the Court promotes the possibility of “moral education” and “moral evolution.”³³³ The Court does not simply reflect our shared constitutional understandings of the past. It also expounds them in light of the old text and the new issues posed in each historical era. To the ancient questions — how shall we live together as a people, and what is most important for us? — the Court’s decisions bear witness to the Bill of Rights in the form of particularized moral and political responses. We should live together so that those who wield the state sovereign sword respect the free expression of our people, their right to be secure in their homes and persons, their right to be free of the establishment of religion, and their right to free exercise of their religious beliefs. In these matters, the Court’s role is to help to define what it means to be a people, a polyglot American people, after two centuries of casting an American form of the human enterprise.

For free exercise, the Court, especially in *Braunfeld*, *Sherbert*, *Thomas*, *Hobbie*, *Frazee*, and *Yoder*, has erected the beginning of a sheltering canopy of respect for the minority religious voices in our midst “yearning to be free” and reminding us of our “better selves.” Those voices long in our midst, including the Native American, the Amish, the Mormon, the Seventh-Day Adventist, and the Jehovah’s Witness, as well as the less familiar voices of the Muslim, the Hindu, the Santabrian, the Shinto, the Buddhist, and the Confucian, can also summon our better selves.

meaning. They saw themselves as being a “people” in the classical and biblical sense of the word. They hoped they were a people of God. They often found themselves to be a people of the devil. . . . Time and again there have arisen prophets to recall this people to its original task, its errand into the wilderness. Significant accomplishments in building a just society have alternated with corruption and despair in America, as in other lands, because the struggle to institutionalize humane values is endless on this earth.

Id. at 98 (quoting R. BELLAH, THE BROKEN COVENANT: AMERICAN CIVIL RELIGION IN A TIME OF TRIAL 2 (1975)).

333. *Id.* at 98-99, 111. “The Supreme Court is, among other things, an educational body, and the Justices are inevitable teachers in a vital national seminar.” *Id.* at 112 (quoting Rostow, *The Democratic Character of Judicial Review*, 66 HARV. L. REV. 193, 208 (1952)). To illustrate, *Brown v. Board of Education*, 347 U.S. 483 (1954), has led to such moral education and moral evolution. Naturally, education works both ways: “[T]he *Plessy* edict led to the expansion of segregation; the *Japanese Exclusion Cases* were relied on to support the McCarran Act’s detention camps; the *Gobitis* decision stimulated flag salute programs. . . .” M. PERRY, *supra* note 329, at 113 (quoting J. CHOPER, JUDICIAL REVIEW AND THE NATIONAL POLITICAL PROCESS: A FUNCTIONAL RE-CONSIDERATION OF THE ROLE OF THE SUPREME COURT 168 (1980)). Closely related is the idea of the constitution and the decisions of the Court as creating a ““rhetorical community’ by which aspects of our culture are defined and redefined.” Nagel, *supra* note 129, at 172.

These heard and unheard voices could invoke shared memories of times past, reminding us that our nation has offered a refuge to prior dissenters who also yearned to be free, including the Pilgrims, the Quakers, the Jews, the Huguenots, and the Catholics, who fled discrimination, expulsion, and death in the old world. In seeking refuge in the new world, all these freedom-seekers yearned for what was refused them in the old world infernos — respect.³³⁴ The first amendment promises respect for voices, old and new, that claim a burdening of free exercise. May the Court be true to its oath to respect this chorus of voices by considering free exercise claims rather than rejecting them *a priori*. In light of our history, may it beacon brightly for the Native American voice.

334. For the importance of what Joel Feinberg calls an attitude of respect "toward the humanity in each man's person," see J. FEINBERG, SOCIAL PHILOSOPHY 91-94 (1973).

Winning Isn't Everything, It's the Only Thing. Violence in Professional Sports: The Need for Federal Regulation and Criminal Sanctions

DANIEL R. KARON*

INTRODUCTION

The National Hockey League (NHL) calls it high-sticking, but if Wayne Maki of the St. Louis Blues had acted as he did toward Ted Green of the Boston Bruins anywhere except on the ice, it would have been called battery. During a September 21, 1969 hockey game, Maki struck Green in the face with his hockey stick. As a result of this attack, Green sustained a serious concussion and massive hemorrhaging and underwent two brain operations that were only partially successful.¹

This is merely one example of the many cases of egregious and excessive violence that take place in professional sports arenas.² Violence has become the rule and no longer the exception in professional sports.³ Of course, some degree of violent contact is necessary in sports, but violence to the degree described above far exceeds this necessary level. Acts that are clearly criminal on the streets seem to be licensed if they take place within the context of a professional sporting event.⁴

The purpose of this Article is to show that the level of violence currently existing in professional sports is intolerable and must be

* B.A., Indiana University-Bloomington; J.D., The Ohio State University College of Law. The author wishes to thank Miriam E. Friedman, B.S.E., Indiana University-Bloomington; M.N.O., Case Western Reserve University for her insight and encouragement in writing this Article.

1. Maki hit Green with his stick in a manner closely resembling a "logger splitting a stump." Green only regained partial sensation and "has never recovered 100 percent." Comment, *It's Not How You Play The Game, It's Whether You Win Or Lose: The Need For Criminal Sanctions To Curb Violence In Professional Sports*, 12 HAMLINE L. REV. 71, 71 n.1 (1988) [hereinafter Comment, *How You Play The Game*]. See also R. HARROW, SPORTS VIOLENCE: THE INTERACTION BETWEEN PRIVATE LAWMAKING AND THE CRIMINAL LAW 19 (1980) (citing *You Hear Green's Name, You Recall That Night*, Boston Globe, Jan. 24, 1979, at 19).

2. See Comment, *How You Play The Game*, *supra* note 1, at 71. See also Note, *Controlling Sports Violence: Too Late for the Carrots—Bring on the Big Stick*, 74 IOWA L. REV. 681 (1989) [hereinafter Note, *Controlling Sports Violence*]; Note, *Sports Violence as Criminal Assault: Development of the Doctrine by Canadian Courts*, 1986 DUKE L.J. 1030 [hereinafter Note, *Criminal Assault*].

3. Comment, *How You Play The Game*, *supra* note 1, at 72.

4. 2 R. BERRY & G. WONG, LAW AND BUSINESS OF THE SPORTS INDUSTRIES 420 (1986) [hereinafter BERRY].

curtailed. Part I explains the societal need for ridding professional sports of needless violence. Part II describes the four major sports leagues' attempts to rid their respective sports of excessive violence. Part III discusses the practical problems and hurdles a prosecutor faces when bringing criminal charges against a professional athlete for an act of on-the-field violence. Part IV describes the federal government's unsuccessful attempts at regulating sports violence through legislation. Finally, Part V proposes a federal statute that will adequately remedy the serious problem of sports violence.

I. THE SOCIETAL NEED FOR RIDING PROFESSIONAL SPORTS OF VIOLENCE

Psychologists and sociologists recognize the problem of sports violence and its negative impact on society.⁵ This destructive societal effect and the effect violence has on the players are the reasons violence must be prohibited.

A. *Professional Athletes as a Role Model for Children*

Athletes serve as role models for our nation's children.⁶ Professional sports have become so pervasive in our society that nearly everywhere one looks, one can expect to see some reference to professional sports.⁷ The "sports star" has been created as a result of this widespread exposure.⁸ Children attempt to emulate those stars who excel in their particular sport.⁹ If the emulated sports stars incorporate violence into their game, a child may display similar violence while playing Little League or while on the playground.¹⁰ A prominent NHL defenseman of the 1970s, Bobby Orr, has even written an instructional book that teaches children that the most effective way to win a hockey game is to fight.¹¹ Professional sports' ambivalent approach toward violence encourages children to adopt the same disinterested attitude.

5. Note, *Controlling Sports Violence*, *supra* note 2, at 686.

6. *Id.* at 688.

7. See Bryant & Zillmann, *Sports Violence and the Media*, in *SPORTS VIOLENCE* 195, 206-07 (J. Goldstein ed. 1983).

8. Note, *Controlling Sports Violence*, *supra* note 2, at 689.

9. Silva, *Factors Related to the Acquisition and Exhibition of Aggressive Sport Behavior*, in *PSYCHOLOGICAL FOUNDATIONS OF SPORT* 261, 269 (J. Silva III & R. Weinberg eds. 1984).

10. Note, *Controlling Sports Violence*, *supra* note 2, at 689 n.74. See Smith, *Significant Others' Influence on the Assimilative Behavior of Young Hockey Players*, 3 INT'L REV. OF SPORT SOC. 45, 54 (1974).

11. B. ORR, *MY GAME* (1974). See Note, *Controlling Sports Violence*, *supra* note 2, at 689 n.78.

B. *Injury to the General Public*

Two principal justifications of our nation's criminal law are that it protects the public and supports notions of morality and ethics.¹² Some argue that criminal law should not apply to acts of on-the-field violence because these acts do not threaten the public in the same way as violence that occurs on the streets.¹³ Thus, criminal law's dual justifications will not be furthered by prosecuting the athletes involved.

This argument, however, fails to address both of the criminal law's justifications. Although the general public is not physically harmed by viewing sports violence, spectators are nevertheless injured by having to view senseless violence—violence that is in direct contravention to the teachings of educational and religious institutions in this country.¹⁴ That the violence occurred within the confines of a playing field or ice rink does not diminish the fact that violence has occurred and that the public has been subjected to viewing it. Accordingly, unless society's notions of morality and ethics endorse violence, these notions have been violated. The argument that sports violence does not implicate the justifications of our nation's criminal law is therefore misleading. Moreover, spectator violence is almost the exclusive result of on-the-field violence.¹⁵ When spectators see their sports heroes act violently on the field, some fans seek to emulate those players.¹⁶

II. LEAGUE ATTEMPTS TO DEAL WITH SPORTS VIOLENCE

A. *League Disciplinary Rules*

All professional sports leagues have established internal rules, systems, and procedures for dealing with violence in their respective sport.¹⁷ These rules exist to avoid the entanglements of civil and criminal proceedings for violent acts that occur during a sporting event.¹⁸ The leagues, however, fail to use their available controls effectively. The

12. BERRY, *supra* note 4, at 420; Note, *Controlling Sports Violence*, *supra* note 2, at 687.

13. BERRY, *supra* note 4, at 420; Note, *Controlling Sports Violence*, *supra* note 2, at 687.

14. Note, *Controlling Sports Violence*, *supra* note 2, at 687.

15. *Id.* See Engler, *Violence in Sport*, in SPORTS IN LAW: CONTEMPORARY ISSUES 180, 181 (H. Appenzeller ed. 1985).

16. Note, *Controlling Sports Violence*, *supra* note 2, at 687-88.

17. BERRY, *supra* note 4, at 433.

18. *Id.*

existing rules do not adequately redress the violence problem, and the leagues have resisted imposing significant rule changes to deal with their current rules' inadequacies.¹⁹

The instruments of league control are players' standard contracts or collective bargaining agreements.²⁰ These agreements frequently contain a clause stating that the player agrees to be bound by the league's disciplinary rules.²¹ Unfortunately, many professional sports contracts and collective bargaining agreements contain nebulous language that authorizes the league commissioner to discipline a player for acting in a manner contrary to the "best interests of the game."²² As a result,

19. See Comment, *How You Play The Game*, *supra* note 1, at 72.

20. BERRY, *supra* note 4, at 433.

21. *Id.*

22. The following are examples of this ambiguous language:

a. National Football League's Standard Player Contract, clause 15: *Integrity of the Game*.

Player recognizes the detriment to the league and professional football that would result from impairment of public confidence in the honest and orderly conduct of NFL games or the integrity and good character of NFL players. Player therefore acknowledges his awareness that if he . . . is [found] guilty of any . . . form of conduct reasonably judged by the League Commissioner to be detrimental to the League of professional football, the Commissioner will have the right, but only after giving Player the opportunity for a hearing at which he may be represented by counsel of his choice, to fine Player in a reasonable amount, to suspend Player for a period certain or indefinitely; and/or to terminate this contract.

Id. at 433 n.1.

b. The Collective Bargaining Agreement between the National Hockey League and the National Hockey League Players Association, art. 15.01, "Intentional Injury," states:

The Clubs shall promulgate on an experimental basis a rule requiring immediate suspension of any hockey player who receives a match penalty for intentionally injuring any other hockey player. Such suspension shall remain in effect until a determination with respect to the match penalty has been made by the President of the National Hockey League.

Id. at 434 n.3.

c. *National Basketball Association's Administrative Manual*, sec. 330 at 2-3: *Position of NBA and Its Teams Regarding Violence on the Basketball Court*:

Violence has no place in the game of basketball and violent behavior cannot be tolerated under any circumstances.

You are highly skilled and highly trained athletes. Your physical well-being is of paramount importance to you, your team, and the NBA. Violence can and often does result in serious physical injury. Such injury could seriously impair your future as an athlete as well as others who may be involved in such violent conduct.

The NBA and all of its teams abhor violence and condemn its existence in our sport. As a player in the NBA, you are hereby advised that violent conduct will not be tolerated under any circumstances. Nothing which occurs

these contracts and agreements fail to take any real or substantial steps toward curbing sports violence.

B. One League's Argument for Its Inactive Stance Toward Violence

The NHL argues that ending hockey violence will cause the league to lose the attendance of those fans who hunger for fights.²³ The league claims that its fans want to see fighting. Moreover, the NHL claims that fighting is simply part of the game, and a rule prohibiting it would be unenforceable.²⁴

Both of these arguments lack merit, however. One of the reasons hockey is marketed to a narrow cross-section of the public is that many people do not care to see fighting or to have their children view it.²⁵ Even the president of the NHL, John Ziegler, does not believe that sports fans attend games merely to see fights.²⁶ It is not surprising that our country's most violent major league sport has no national network television contract. If fighting and excessive violence are removed from hockey, its appeal may actually broaden.²⁷ Nevertheless, the NHL, for whatever reason, refuses to take a more active stance toward the violence occurring in its league.

III. BRINGING CRIMINAL CHARGES AGAINST ATHLETES

A. Prosecutors' Unique Perspective on Sports Violence

When violence and criminal behavior occur on the sports field, prosecutors view this behavior differently than if it had occurred on the street. Understanding the basis of criminal law makes it easier to understand why this is the case. Recall that one of the criminal law's

during a game can justify an act of willful violence.

The NBA and your team will take immediate and appropriate action against any player who engages in such conduct and all personnel are advised that violence must be avoided at all times. There will be no variance from this express statement of policy.

This avoidance of violence is to your benefit, as well as to the benefit of all players and teams in the NBA. You must comply.

Id. at 434 n.4 (emphasis in original).

23. *End Hockey Violence Now*, Christian Science Monitor, Apr. 16, 1986, at 14, col. 3 (opinion).

24. *Id.*

25. *Id.*

26. *Hockey and Fighting: Uneasy Coexistence*, Washington Post, Jan. 26, 1989, § C, at 1, col. 2.

27. *Id.*

dual justifications is that it protects the public.²⁸ When an incident of sports violence occurs, the harm is confined to the game's participants.²⁹ These participants know and assume the risks of the game.³⁰ Furthermore, the public is not subjected to any risk of harm because spectators are either in the stands or at home watching the sports event on television.³¹ Sports participants know what they are "getting themselves into," and only they will be physically harmed by other players' behavior. Therefore, prosecutors are reluctant to view sports violence as behavior worthy of their attention.

B. *Defenses Raised When a Prosecution for Sports Violence is Brought*

1. *Defining Criminal Conduct*.—When a prosecutor brings criminal charges against an athlete for behaving violently on the field, the charge is typically battery.³² Criminal battery can be defined briefly as "the unlawful application of force to the person of another."³³ The requirement that a criminal battery be "unlawful" is the key to understanding why sports violence is often treated as noncriminal.³⁴ Society and prosecutors seem to treat sports violence as "lawful" behavior.³⁵ Consequently, an act that is criminal on the streets becomes legal on the field because the "unlawful" requirement is negated.³⁶ Sometimes, however, on-the-field behavior is so heinous that it exceeds the level of contact that is considered lawful within the rules of a game.³⁷ The problem is drawing the line between "lawful" conduct, that is within the rules of a particular game (including accidental contact), and conduct that is criminal.³⁸

2. *Consent*.—A common defense raised by athletes is that the victim consented to the violent contact.³⁹ Normally, consent is not a defense to criminal battery because a criminal offense is a wrong that affects the general public, at least indirectly.⁴⁰ Consequently, the actor cannot

28. BERRY, *supra* note 4, at 420.

29. *Id.*

30. *Id.*

31. *Id.*

32. *Id.* at 422.

33. W. LAFAYE & A. SCOTT, CRIMINAL LAW 685 (1986) [hereinafter LAFAYE].

34. BERRY, *supra* note 4, at 422.

35. *Id.*

36. *Id.*

37. *Id.*

38. *Id.*

39. *Regina v. Ciccarelli*, No. 2388 (Ontario Dist. Ct.-York Jud'l Dist., Dec. 21, 1989) (unpublished decision) (on file in the office of the Indiana Law Review).

40. BERRY, *supra* note 4, at 423; LAFAYE, *supra* note 33, at 477; J. WEISTART & C. LOWELL, THE LAW OF SPORTS 185 (1979) [hereinafter WEISTART].

be licensed by the person directly harmed.⁴¹ Put another way, the public interest may not be frustrated by private license.⁴² With respect to sports violence, however, consent is a valid defense. The Model Penal Code states that consent may be a defense to criminal charges of battery arising from conduct in a sports event.⁴³

An athlete is understood to impliedly consent to a certain amount of physical contact on the field or rink.⁴⁴ The problem is measuring the level of violence to which a player impliedly consents.⁴⁵ The Model Penal Code states that an athlete consents to a "reasonably foreseeable" amount of hazard or violence in sport.⁴⁶ With respect to the Model Penal Code, the difficult issue is distinguishing between reasonably foreseeable hazards that are consented to as part of the sport and hazards for which there is no consent.⁴⁷

One approach to distinguishing between behavior that is consented to and behavior for which there is no consent concerns conduct that is customarily part of a particular sport.⁴⁸ Under this approach, a player will be deemed to consent to conduct that is normally associated with the particular sport in which the player participates.⁴⁹ This approach, however, is too broad.⁵⁰ Fisticuffs and stick fighting, for example, are customary activities in professional hockey. Under this approach, a player will be understood to consent to these violent activities.⁵¹

An alternative approach is the "rules-of-the-game" test.⁵² Under this test, a player will not be deemed to consent to acts that are illegal under the rules of the sport.⁵³ This approach is too narrow in scope.⁵⁴

41. LAFAVE, *supra* note 33, at 477.

42. WEISTART, *supra* note 40, at 185.

43. The Model Penal Code, § 2.11(2) (Proposed Official Draft 1962) provides: [W]hen conduct is charged to constitute an offense because it causes or threatens bodily injury, consent to such conduct or to the infliction of such injury is a defense if: . . . (b) the conduct and the injury are reasonably foreseeable hazards of joint participation in a lawful athletic contest or competitive sport.

44. *Regina v. Ciccarelli*, No. 2388 (Ontario Dist. Ct.-York Jud'l Dist., Dec. 21, 1989) (unpublished decision) (on file in the office of the Indiana Law Review); Note, *Criminal Assault*, *supra* note 2, at 1038.

45. *Regina v. Ciccarelli*, No. 2388 (Ontario Dist. Ct.-York Jud'l Dist., Dec. 21, 1989) (unpublished decision) (on file in the office of the Indiana Law Review).

46. MODEL PENAL CODE § 2.11(2) (Proposed Official Draft 1962).

47. BERRY, *supra* note 4, at 427.

48. *Id.*; WEISTART, *supra* note 40, at 186.

49. BERRY, *supra* note 4, at 427.

50. WEISTART, *supra* note 40, at 186.

51. *Id.*

52. BERRY, *supra* note 4, at 427.

53. *Id.*

54. WEISTART, *supra* note 40, at 186.

This test considers all technical violations of a sport to be conduct to which the player did not consent. For example, intentional fouls in basketball, committed for strategic purposes, will be considered conduct to which there is no consent and criminal liability could result.⁵⁵

A third approach looks to the seriousness of the injury.⁵⁶ Under this approach, a victim is deemed not to have consented to behavior that causes grave injury.⁵⁷ The problem with this final approach is that a player does not know beforehand whether contact with the victim will result in serious injury. The most innocent hit, if the recipient lands poorly, for example, could result in grave injury. Consequently, this test does not establish an adequate means by which a player can gauge his conduct.

No useful American authority exists that clarifies the consent issue. Canadian cases, however, provide some guidance,⁵⁸ and when coupled with the Model Penal Code's "reasonably foreseeable" approach, one can derive the following standard: "An athlete will not be deemed to have consented to intentional or reckless acts that are not reasonably related to the conduct of the sport in question."⁵⁹ This standard, however, raises the amorphous issue of "reasonableness" and from whose standpoint it is to be considered.

Consent then, is an evasive concept with respect to the prosecution of a professional athlete. This concept's unstable character, however, does not mean that consent is not readily used as a defense. Consent is a valid and occasionally successful defense for an athlete.

3. *Self-Defense*.—A second defense that may be asserted by an athlete is self-defense. One description of self-defense is:

One who is not the aggressor in an encounter is justified in using a reasonable amount of force against his adversary when he reasonably believes (a) that he is in immediate danger of unlawful bodily harm from his adversary and (b) that the use of such force is necessary to avoid this danger.⁶⁰

In sports, however, this defense may not be available for a number of reasons.⁶¹

First, the athlete must show that he used only reasonable force in repelling an attack.⁶² This issue will usually arise in cases of escalating

55. *Id.*

56. BERRY, *supra* note 4, at 427.

57. *Id.*

58. Regina v. Bradshaw, 14 Cox Crim. Cas. 83 (Leicester Spring Assizes 1878).

59. WEISTART, *supra* note 40, at 188.

60. LAFAVE, *supra* note 33, at 454.

61. BERRY, *supra* note 4, at 427.

62. *Id.* at 454.

violence, as when one hockey player punches an opponent and the opponent strikes back with his stick.⁶³ In this situation, the opponent (now on trial) may have difficulty establishing that he used only reasonable force (*i.e.*, no greater than the force used against him) in repelling the aggressor's attack.⁶⁴

Next, the defendant must have honestly believed that the danger of immediate, serious bodily injury was imminent.⁶⁵ Often, the defendant does not have this honest belief because he provoked the attack.⁶⁶ Thus, this element of the defense is negated.⁶⁷

Finally, this defense is limited to those cases in which the defendant had no reasonable means of retreat and cases in which force was necessary to avoid danger.⁶⁸ If a player could have stopped the confrontation by retreating, he cannot plead self-defense. Often, a player can end the engagement by simply stepping back, skating across the ice, or allowing a referee or umpire to intervene. Self-defense, therefore, is not a useful theory because many times one or all of its elements are defeated by the defendant's actions.

C. *Prosecutorial Hesitancy to Bring Criminal Charges Against Athletes*

Although prosecutors may, in their discretion, bring criminal charges against professional athletes, they are rarely motivated to bring these charges. Moreover, if charges are brought, a prosecutor may have difficulty obtaining a guilty verdict.

1. *Prosecutorial Restraint.*—Prosecutors are generally reluctant to bring criminal charges against sports figures. The injured player often does not want to file criminal charges.⁶⁹ As a result, the prosecutor does not have a complainant. Moreover, the prosecutor has a "poor" witness if there is a witness at all.⁷⁰ Players are often reluctant to testify because they either subscribe to the "code" of the particular sport, or they are ambivalent toward violence.⁷¹

Furthermore, prosecutors are concerned about the fairness of a sports figure's trial because of juries' tendencies to support the home-

63. *Id.*

64. *Id.*

65. *Id.*; LAFAVE, *supra* note 33, at 454.

66. BERRY, *supra* note 4, at 427; Comment, *How You Play the Game*, *supra* note 1, at 85.

67. BERRY, *supra* note 4, at 427.

68. *Id.*; LAFAVE, *supra* note 33, at 454.

69. *Violence in Sports: Specialists See Obstacles to Bringing Athletic Fights to Court*, Boston Globe, Jan. 13, 1987, at 1, col. 3.

70. *Id.*

71. *Id.* at 3.

town team.⁷² For example, some speculate that anyone who fights with Larry Bird during an NBA game in Boston will be convicted by a Boston jury. On the other hand, some predict that Bird would most likely be acquitted by this same jury if he was involved in the altercation.⁷³

Next, prosecutors rarely bring charges because of their own beliefs that league rules and sanctions are sufficient to curb sports violence.⁷⁴ Prosecutors are sports fans, and it is not surprising that they too subscribe to the erroneous notion that leagues can adequately police themselves. Prosecutors also believe that the prosecution of athletes will be too time consuming. If prosecutors decided to bring charges against athletes, police would have to arrest these players immediately.⁷⁵ Consequently, police would be assigned to every sporting event.⁷⁶ Prosecutors believe that "criminal prosecution has little or no role in controlling [sports] violence when it is confined to game time."⁷⁷ Prosecutors assert that they have "enough straight line criminal violence to keep [themselves] busy without entering a new media game."⁷⁸

Professor Wayne R. LaFave believes prosecutors are reluctant to bring criminal charges against athletes for two reasons. The first is the "community subgroup rationale."⁷⁹ According to this rationale, if certain illegal activity is pervasive in a particular subgroup of society, this activity should be tolerated.⁸⁰ LaFave's second rationale concerns the crime's setting and whether the purposes of criminal law will be furthered by prosecuting the actor. He argues that because law in the area of sports violence is basically ineffective and inefficient, prosecutors simply choose not to enforce it.⁸¹

A final reason that the prosecution of athletes may be disfavored by prosecutors is the "continuing relationship" theory.⁸² This theory

72. *Id.*

73. *Id.*

74. Sprotzer, *Violence in Professional Sports: A Need for Federal Regulation*, 86 CASE & COM. 3, 6 (1981).

75. *Political Penalty Box*, Boston Globe, Jan. 8, 1987, at 12, col. 1.

76. *Id.*

77. R. HARROW, *supra* note 1, at 114.

78. *Id.*

79. W. LAFAVE, *ARREST: THE DECISION TO TAKE A SUSPECT INTO CUSTODY* 110 (1965).

80. *Id.*

81. *Id.*

82. Comment, *Assumption of Risk and Vicarious Liability in Personal Injury Actions Brought by Professional Athletes*, 1980 DUKE L.J. 742, 753 [hereinafter Comment, *Assumption of Risk*].

positives that because only athletes are physically injured by sports violence, the law may be more tolerant of this violence.⁸³

2. *Difficulty in Obtaining a Guilty Verdict.*—If a prosecutor decides to bring criminal charges against an athlete, resistance from the jury may be encountered. Juries are reluctant to convict athletes for on-the-field violence.⁸⁴ For example, on January 4, 1974, during an NHL game in Minnesota, the Boston Bruins' Dave Forbes permanently injured Henry Boucha of the Minnesota North Stars.⁸⁵ Forbes punched Boucha in the face with the hand that held his hockey stick. The butt end of the stick struck Boucha in the right eye.⁸⁶ Boucha fell to the ice, and Forbes pounced on him and pummeled his head into the ice.⁸⁷

Forbes was indicted by a Minneapolis grand jury and charged with aggravated assault by use of a dangerous weapon.⁸⁸ Unfortunately, a week and a half of trial testimony resulted in a hung jury.⁸⁹ Interviews with the twelve jurors revealed a nine to three split in favor of conviction.⁹⁰ The three jurors who voted for acquittal believed that the hockey community accepted fighting as part of the game.⁹¹ One of these three jurors commented that he did not believe it was proper to make Forbes the "scapegoat" for the problems of the entire sport.⁹²

IV. ATTEMPTS AT FEDERAL REGULATION

A. *The Sports Violence Act of 1980*

In 1980, Ohio Representative Ronald M. Mottl introduced the Sports Violence Act (1980 Act) into the House of Representatives.⁹³ Proponents

83. *Id.*

84. State v. Forbes, No. 63,280 (Hennepin Cty. Minn. Dist. Ct.) (dismissed Aug. 12, 1975). See Flakne & Caplan, *Sports Violence and the Prosecution*, 13 TRIAL 33, 34 (1977) (written by two prosecutors involved in the *Forbes* case); Note, *Controlling Sports Violence*, *supra* note 2, at 701.

85. State v. Forbes, No. 63,280 (Hennepin Cty. Minn. Dist. Ct.) (dismissed Aug. 12, 1975); Note, *Controlling Sports Violence*, *supra* note 2, at 701.

86. Note, *Controlling Sports Violence*, *supra* note 2, at 701.

87. *Id.*

88. *Id.* at 701-02.

89. *Id.* at 702.

90. *Id.*

91. *Id.*

92. *Id.* This juror stated, "Just like if I had committed some crime because of my job then my employer should suffer or should answer [for] it—not me." Hallowell & Meshbesher, *Sports Violence and the Criminal Law*, 13 TRIAL 27, 28 (1977) (the authors of this article were Forbes' defense attorneys).

93. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980) (this bill was reintroduced by Jack Kemp (R. N.Y.), former quarterback for the Buffalo Bills, in 1981 as H.R. 2263); Sprotzer, *supra* note 74, at 4 (the 1980 bill was co-authored by Richard B. Morrow, a Harvard Law School graduate).

of the 1980 Act claimed that federal regulation is necessary because existing league mechanisms fail to curb violence effectively.⁹⁴ They also argued that league self-regulation does not deter athletes from participating in violent behavior, and state and local criminal laws are not enforced.⁹⁵ The bill never made it onto the floor of the House.⁹⁶

The bill's early death was not a tragedy because it would have created more problems than it would have solved.⁹⁷ The bill would have imposed criminal liability on players who use "excessive physical force."⁹⁸ "Excessive physical force" was defined in subsection (b) of the bill as physical force that "(A) ha[d] no reasonable relationship to the competitive goals of the sport; (B) was unreasonably violent; and (C) could not be reasonably foreseen, or was not consented to, by the injured person, as a normal hazard of such person's involvement in such sports event."⁹⁹ For behavior to be sanctionable, it would have to meet the three criteria set out in this subsection.

This bill's primary problem was that subsection (b) was vague.¹⁰⁰ In defining "excessive violence," subsection (A) referred to the "competitive goals of the sport."¹⁰¹ This subsection, however, did not specifically define the competitive goals of any sport.¹⁰² Arguably, the parties involved in any sport have varying goals.¹⁰³ For players, the goal is winning.¹⁰⁴ The goal of team owners is profit, which is directly related to a team's success.¹⁰⁵ Finally, league officials are concerned with the entertainment value of their sport.¹⁰⁶ Clearly, the competitive goal to which subsection (A) referred had to be defined before it could be determined that contact constituted "excessive physical force" under the Act.¹⁰⁷

In addition to its vagueness, this subsection did not further the goals of the proposed enactment. For example, if one could show that exhibitions of brutal force furthered the various goals outlined above

94. Comment, *Assumption of Risk*, *supra* note 82, at 1032.

95. *Id.* at 1032-33.

96. See Comment, *How You Play The Game*, *supra* note 1, at 79.

97. Note, *Controlling Sports Violence*, *supra* note 2, at 691.

98. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980).

99. *Id.*

100. Note, *Controlling Sports Violence*, *supra* note 2, at 691.

101. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980).

102. Note, *Controlling Sports Violence*, *supra* note 2, at 691.

103. *Id.*

104. *Id.*

105. *Id.*

106. *Id.* at 692.

107. *Id.*

(for example, the “rougher” a team is, the more likely it will win a game), this team behavior could be considered not “excessively violent.” Obviously, this was not the intended result of the bill.

Subsection (B) defined “excessive physical force” as force that is “unreasonably violent.”¹⁰⁸ This is a circular definition.¹⁰⁹ Determining whether conduct is “unreasonably violent” does not help to determine whether it is excessively violent.¹¹⁰ Moreover, the bill failed to define the term “unreasonable” as applied to a professional athlete’s behavior.¹¹¹ Whether the reasonableness of an athlete’s conduct should be compared to that of an average person or an average player is unclear.¹¹² Assuming this standard relates to the average sports player and the average player commonly engages in violent behavior, an athlete prosecuted under the 1980 Act would have been acting reasonably if he too had acted violently. Again, this could not have been the intended result, yet subsection (B) could lead to this conclusion.

The third requirement for “excessive physical force” was contained in subsection (C). This subsection concerned lack of reasonable foreseeability and consent.¹¹³ Again, the 1980 Act failed to adequately state from whose perspective these terms were to be determined.¹¹⁴ For example, an NHL “enforcer”¹¹⁵ may reasonably foresee or consent to violent treatment on the ice, but a less physical player, like Wayne Gretzky, might not expect or consent to brutal behavior.

Subsection (C)’s reference to contact exceeding a game’s “normal hazards” only exacerbated this problem.¹¹⁶ According to this portion of the subsection, if contact is not reasonably foreseen or consented to as a “normal hazard” of a game, it is excessively violent. Determining whether an activity is a normal hazard, however, depends on the current practices within a sport. If a certain game sanctions an unusual amount of violence (for example, hockey’s endorsement of fighting), this violence may be considered a normal hazard of the game.¹¹⁷ Because all normal hazards of a sport are, by nature of their normality, “reasonably foreseeable” by-products of that sport, frequently occurring violent acts would be permissible under subsection (C).

108. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980).

109. Note, *Controlling Sports Violence*, *supra* note 2, at 692.

110. *Id.*

111. *Id.*

112. *Id.*

113. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980).

114. Note, *Controlling Sports Violence*, *supra* note 2, at 693.

115. An “enforcer” is a player whose job is to protect the team’s star player from overzealous opposing players.

116. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980).

117. Note, *Controlling Sports Violence*, *supra* note 2, at 693.

Finally, the 1980 Act was too narrow in scope.¹¹⁸ The Act defined a “professional sports event”¹¹⁹ as a “paid-admission contest.”¹²⁰ This definition is ambiguous because it does not clarify whether free collegiate events are included in the Act’s regulation.¹²¹ Moreover, whether the Act would have conferred professional status on collegiate sporting events that charge admission (such as NCAA basketball and football contests) is unclear.¹²²

B. *The Sports Violence Arbitration Act of 1983*

The Sports Violence Act of 1980 was fraught with problems and would have been unworkable. In 1983, South Dakota Representative Thomas A. Daschle sought to introduce another bill into the House of Representatives that addressed the problem of sports violence.¹²³ Unlike the 1980 Act, the Sports Violence Arbitration Act of 1983 (1983 Act) would have imposed civil, rather than criminal liability on athletes.¹²⁴ This bill also failed.¹²⁵

The 1983 Act proposed the creation of an arbitration board to assist players in settling grievances resulting from “conduct found to be inconsistent with the competitive goals of [the] sport.”¹²⁶ The bill called for management and players to create a board and to establish procedures through collective bargaining.¹²⁷

The 1983 Act, however, like its 1980 predecessor, had inherent problems. The bill failed to recognize the leagues’ internal resistance to outside attempts at forced self-regulation.¹²⁸ Players and leagues probably would not have taken the initiative required to establish and implement the necessary arbitration boards.¹²⁹ Moreover, players would have been disinclined to bring grievances because of their belief that revenge is the most effective and satisfying way to deal with on-the-field attacks.¹³⁰

118. Sprotzer, *supra* note 74, at 8.

119. H.R. 7903, 96th Cong., 2d Sess., 126 CONG. REC. 20,890 (1980).

120. *Id.*

121. Sprotzer, *supra* note 74, at 10.

122. Note, *Controlling Sports Violence*, *supra* note 2, at 693.

123. H.R. 4495, 98th Cong., 1st Sess., 129 CONG. REC. H10,579 (1983). This act was modeled after Note, *The Sports Court: A Private System to Deter Violence in Professional Sports*, 55 S. CAL. L. REV. 399 (1982).

124. H.R. 4495, 98th Cong., 1st Sess., 129 CONG. REC. H10,579 (1983).

125. BERRY, *supra* note 4, at 435; Comment, *How You Play The Game*, *supra* note 1, at 79.

126. H.R. 4495, 98th Cong., 1st Sess., 129 CONG. REC. H10,579, § 3(2)(A) (1983).

127. *Id.*

128. Note, *Controlling Sports Violence*, *supra* note 2, at 693.

129. *Id.* at 694.

130. *Id.*

V. A WORKABLE FEDERAL STANDARD

A. *Restatement of the Problem*

A workable federal standard that provides criminal sanctions for egregious acts of sports violence is needed. State and local laws are not remedying the sports violence problem because of prosecutorial hesitancy, and federal attempts have failed to solve the problem. A well-drafted federal standard will provide consistent treatment of athletes and will mandate the prosecution of violent acts. With a specific federal statute, prosecutors will not hesitate to bring charges against athletes.

The problem previous legislators had in drafting an adequate statute was their inability to articulate the idea that intentionally inflicted, excessive force exhibited by an athlete should be punished. In attempting to articulate this idea, the legislators used terminology that caused conceptual difficulties. The key issue is to adequately address and define the type and level of violence that will give rise to criminal liability.

B. *A Proposed Federal Statute*

The following act is a workable way to rid all sports of intolerable violence: "Athletes will be held in violation of this Act, and will be subject to criminal penalties thereunder, for intentionally inflicting contact meant to cause physical injury to an opposing player, as opposed to contact intended to further the goals of the particular competition."

1. "*Athlete.*"—Under the proposed act, the term "athlete" will mean any athlete, whether professional or amateur, engaged in any sporting event or competition. This definition eliminates the problems of scope and answers any questions regarding to whom or to what events the act applies.

2. *The First "Intent" Requirement.*—The act contains an intent requirement regarding contact. If the contact is accidental, no criminal liability arises. At this point, the reader will be referred to examples of intentional contact that illustrate exactly what contact is being implicated. For example, with respect to hockey, intentional contact is striking an opposing player with one's stick or engaging in fisticuffs. Turning to football, an example of intentional contact is forceful contact with a player that is unrelated to the intended play or that occurs after the play has ended, unless it is accidental. (Accidental contact is contact that is the result of a player's inability to stop his own momentum that forces him into the contacted player). Another example is slamming a player to the ground, as opposed to merely tackling him. With respect to baseball, examples of intentional conduct are

“brushback” pitches or “beanballs” when a pitcher is ordered to pitch them. The particular circumstances existing in a game (*e.g.*, the pitcher’s team is trailing by a wide margin or the hitter has already hit a home run off the pitcher) should be considered in determining the requisite level of intent. These few examples are not intended to constitute an exhaustive list of the types of intentional contact that can take place in a sporting event. They are merely intended to assist in distinguishing intentional contact from accidental, incidental, or inadvertent contact.

3. The Second “Intent” Requirement.—The contact must be further intended to specifically cause physical injury to the opposing player. Without this additional, or secondary, intent requirement, all contact that was intended will be punishable even if it was within the rules of a particular sport or was used for intimidation or strategic purposes. Although any intentional contact may be criminal outside the sports field or rink, this Act’s secondary intent requirement recognizes the inherent nature of contact in sports. The enactment does not, however, recognize this necessary contact to the extent that intentional contact, if meant to injure, will go unpunished. This intent element, along with the first intent element, will be determined by a jury according to the facts and the witnesses’ respective credibilities.

4. Defining “Goals of the Competition.”—The final part of the proposed act relates to contact that is intended to further the goals of the competition and is intended to distinguish punishable contact from permissible contact. Under the act, the term “goals of the competition” is defined from the players’ perspective as winning the game. This last clause helps one further understand that contact that is intended to help a team win is permissible if it is not the type outlined above. Examples will again be given at this point. Permissible contact will include tackling, checking, blocking, and “jockeying” for position. Again, this list is not intended to be exclusive or exhaustive, but merely illustrative.

5. Avoidance of Ambiguities.—The proposed act stays clear of the ambiguities and vague references that doomed the 1980 proposal. Illustrations and examples will be furnished to guide the reader in understanding the intended meaning of the act. These illustrations will prevent the reader from straying into uncertain and obscure analyses with respect to the purpose of the statute. Furthermore, the act does not mention the terms “violent,” “excessive,” “foreseeable,” “consent,” or “reasonable.” These are words that led to harsh criticism of the 1980 Act.¹³¹

131. For criticism of the 1980 Act, see Sprotzer, *supra* note 74, at 8-10; Note, *Controlling Sports Violence*, *supra* note 2, at 693.

VI. CONCLUSION

League self-regulation has not deterred violent acts in the sports arena, and relevant state and local laws are not being enforced. Our nation's criminal laws must finally be applied uniformly to sports conduct. Nothing inherent in the nature of sport allows it to remain insulated from the criminal sanctions applicable to the remainder of society.¹³² The sports arena should not be a sanctuary for unbridled violence to which the criminal laws of our nation do not apply.¹³³ A federal statute of the type proposed will preserve the vitality of sports while serving notice to players that they no longer have license to commit unwarranted batteries on fellow players.¹³⁴ With an effective federal statute, the public will no longer be forced to view acts of unnecessary and senseless violence, and athletes will finally learn the meaning of sportsmanship.

132. WEISTART, *supra* note 40, at 185.

133. *Regina v. Ciccarelli*, No. 2547 (Ontario Prov. Ct.-Crim. Div., Aug. 24, 1988), *aff'd*, No. 2388 (Ontario Dist. Ct.-York Jud'l Dist., Dec. 21, 1989) (unpublished decision) (on file in the office of the Indiana Law Review).

134. Sprotzer, *supra* note 74, at 10.

Indiana Law Review

Volume 25

1991

Number 1

NOTES

CERCLA's Web of Liability Ensnares Secured Lenders: The Scope and Application of CERCLA's Security Interest Exception

INTRODUCTION

In the wake of the twentieth anniversary of Earth Day and the "greening" of America, the American public has become aware of the threat to its health¹ and environment caused by the unsafe disposal of hazardous wastes.² In the zeal to correct society's numerous environmental problems, secured lenders are increasingly incurring liability for clean-up costs associated with their borrowers' contaminated property. In the past, secured lenders could avoid environmental liability by not participating in the management of a borrower's daily operations and by not acquiring title to contaminated property upon foreclosure. Recently, however, the Eleventh Circuit Court of Appeals delineated a new liability trap by ruling that secured lenders may incur liability by merely exhibiting the capacity to influence a borrower's hazardous waste disposal decisions.³ This decision has been challenged by other courts,⁴ and it is now unclear

1. Note, *Toxic Waste Litigation, Developments in the Law*, 99 HARV. L. REV. 1458, 1462 (1986) [hereinafter Note, *Toxic Waste Litigation*]. "Exposure to hazardous wastes can cause cancer, genetic mutation, birth defects, miscarriages, and damage to lungs, liver, kidneys, or nervous system." *Id.* "Evidence has established a causal link between toxic chemical exposure and . . . personality disorders." Burkhart, *Lender/Owner's and CERCLA: Title and Liability*, 25 HARV. J. ON LEGIS. 317, 317 (1988).

2. This Note uses the terms "hazardous waste," "hazardous substance," and "toxic waste" to mean "hazardous substance" according to the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA), Pub. L. No. 96-510, 94 Stat. 2767 (1980) (codified in part as amended at 42 U.S.C. §§ 9601-75 (1988)). See *infra* note 27 and accompanying text.

3. United States v. Fleet Factors Corp., 901 F.2d 1550 (11th Cir. 1990).

4. See *infra* note 89.

what actions by secured lenders will lead to environmental liability.

Although American industry has generated hazardous substances as products and by-products of manufacturing for years, the amount of hazardous substances generated has increased tremendously. American industry generated approximately 150 million metric tons of hazardous waste in 1981. This figure rose to an estimated 266 million metric tons per year by 1986.⁵ This increase is not surprising because the Environmental Protection Agency (EPA) has identified more than 4,000 types of businesses and industries that generate hazardous wastes.⁶

As the amount of hazardous waste generated in this country has increased, so have the costs involved with its clean-up and proper disposal.⁷ The United States General Accounting Office has acknowledged that there are as many as 425,380 potential Superfund sites in the United States.⁸ The estimated response costs for the worst 2,500 of these could amount to more than \$22 billion,⁹ while the average clean-up costs for the other sites could range from \$10-\$12 million per site.¹⁰

In 1980, as a response to public concern over the problem of improper hazardous waste disposal, Congress hastily¹¹ passed the Comprehensive

5. Comment, *Lender Liability for Hazardous Waste Cleanup*, 1988 Wis. L. REV. 139, 140. Most hazardous substances are not destroyed, but are stored perpetually at hazardous waste sites. The process of storing hazardous wastes consists of sealing the waste in drums and burying it in clay-lined dumps. Of the hazardous dumps in existence, 74% use containers, 54% use tanks, 17% use surface impoundments, 6% incinerate wastes, and 5% use landfills. Note, *Toxic Waste Litigation*, *supra* note 1, at 1462.

6. See Burkhart, *supra* note 1, at 320 n.4. The EPA's list of sources potentially generating hazardous waste includes: automobiles; banking, aircraft, aerospace, communications, and public utility industries; electronic, furniture, textile, food, beverage, and grocery manufacturers; and optical, paper, packing, and rubber products.

7. For example, nearly 230 families were forced to evacuate from Love Canal, New York because their homes were built around an abandoned chemical dump containing 350 million pounds of industrial waste. H.R. REP. NO. 1016, 96th Cong., 2d Sess., pt. 1, reprinted in 1980 U.S. CODE CONG. & ADMIN. NEWS 6119, 6121-22. By 1980, more than \$27 million had been spent in response costs. *Id.* Union Carbide Corporation recently agreed to pay at least \$40 million for clean-up of a Colorado hazardous waste site. Wall St. J., Nov. 3, 1986, at 47, col. 1. The EPA has estimated the response costs of the Stringfellow Acid Pits hazardous waste site to be \$40 million. Wall St. J., Sept. 25, 1986, at 23, col. 1. See also Comment, *supra* note 5, at 140-41 nn.3 & 10. The United States government also intends to spend \$33 million to buy back all of the homes in Times Beach, Missouri, a town contaminated by dioxin sprayed on its streets a decade earlier. See Burkhart, *supra* note 1, at 318 n.2.

8. See 18 Env't Rep. (BNA) 2043 (Jan. 22, 1988).

9. *Id.*

10. Burkhart, *supra* note 1, at 318 n.3.

11. CERCLA was adopted during the final days of the 96th Congress just prior to the inauguration of the Reagan administration. The legislation was adopted under a special suspension of rules which precluded any amendments. There was not even a

Environmental, Response, Compensation and Liability Act (CERCLA).¹² In 1986 Congress also passed the Superfund Amendments and Reauthorization Act (SARA).¹³ CERCLA, unlike previous environmental legislation,¹⁴ vested federal and state governments with the authority to respond promptly to releases¹⁵ or threatened releases of hazardous substances from sites existing both prior to and after the statute's enactment.¹⁶ CERCLA authorizes the EPA to target specific waste sites across the nation and to rank the sites through a National Priority List (NPL) which determines the order in which the sites will be cleaned up.¹⁷

CERCLA initially created a trust fund, often referred to as the "Superfund," which provides funds to be used specifically for the clean-up of hazardous waste sites on the NPL.¹⁸ Although the primary responsibility for utilizing Superfund finances for clean-up operations rests with the EPA, state and local governments may also tap the Superfund to finance their own clean-up efforts.¹⁹

conference held on the measure, and no report was issued on the legislation as enacted. For a discussion of CERCLA's legislative history, see Grad, *A Legislative History of the Comprehensive Environmental Response, Compensation and Liability ("Superfund") Act of 1980*, 8 COLUM. J. ENVTL. L. 1 (1982).

12. 42 U.S.C. §§ 9601-75 (1988).

13. In October 1986, Congress passed SARA, which extended the funding for the trust fund for an additional five years. SARA also increased CERCLA funding from \$1.6 billion for the first five years to \$8.5 billion for the five years following SARA's enactment. *Id.* § 9611.

14. Congress initially attempted to address the environmental problem by enacting the Resource, Conservation and Recovery Act (RCRA), which regulates hazardous waste from its generation to its disposal. Pub. L. No. 94-580, 90 Stat. 2795 (1976) (codified as amended at 42 U.S.C. §§ 6901-87 (1988)). RCRA has some deficiencies, however, because it "is of no help if a financially responsible owner of the site cannot be located." H.R. REP. NO. 1016, 96th Cong., 2d Sess., pt. 1, *reprinted in* 1980 U.S. CODE CONG. & ADMIN. NEWS 6119, 6125. Furthermore, RCRA does not provide for site clean-up "when the owner is unknown, is not responsible, or is financially unable to pay for these costs." *Id.* Finally, RCRA "applies to past sites only to the extent that they are posing an imminent hazard." *Id.* See also Note, *The Toxic Mortgage, CERCLA Seeps into the Commercial Lending Industry*, 63 ST. JOHN'S L. REV. 839, 841-43 (1989).

15. The term "release" is generally defined in CERCLA as "any spilling, leaking, pumping, pouring, emitting, emptying, discharging, injecting, escaping, leaching, dumping, or disposing into the environment (including the abandonment or discarding of barrels, containers, and other closed receptacles containing any hazardous substances or pollutant or contaminant)." 42 U.S.C. § 9601(22) (1988).

16. CERCLA applies retroactively to all responsible parties and hazardous waste sites. See *infra* notes 54-55 and accompanying text.

17. 42 U.S.C. § 9605(c) (1988).

18. Pub. L. No. 96-510, § 221, 94 Stat. 2801 (1980), *repealed by* Pub. L. No. 99-499, Title V, § 517(c)(1), 100 Stat. 1774 (1986). Currently, the fund is labeled the "Hazardous Substance Superfund." 42 U.S.C. § 9601(11) (1988).

19. 42 U.S.C. § 9604(d) (1988).

CERCLA provides three mechanisms to initiate the clean-up of hazardous waste that responsible parties fail to clean-up. First, the EPA is empowered to issue an administrative order directing responsible parties to engage in clean-up efforts.²⁰ To enforce these orders, the EPA has the authority to levy fines up to \$25,000 a day.²¹ Second, the EPA can request an injunction in federal court requiring the responsible party to clean-up or abate the release of hazardous substances.²² Third, the government may choose to clean-up the site itself using Superfund money.²³ Under this remedy, CERCLA requires the EPA to sue the responsible parties to reimburse the fund.²⁴ To recover its response costs in such a suit, the government must prove that:

- (1) the site is a "facility"²⁵ for purposes of CERCLA;
- (2) a release²⁶ or threatened release of any hazardous substance²⁷ from the site has occurred;
- (3) the release or threatened release has caused the federal government to incur response costs; and
- (4) the defendant is one of the persons designated as a party liable for costs under CERCLA.²⁸

When CERCLA was enacted, Congress attempted to place the responsibility for hazardous waste clean-up on all parties even remotely

20. *Id.* § 9606(a), (b).

21. *Id.*

22. *Id.* § 9606(a).

23. *Id.* § 9604.

24. *Id.* § 9607(a).

25. The term "facility" means "any building, structure, installation, equipment, pipe or pipeline (including any pipe into a sewer or publicly owned treatment works), well, pit, pond, lagoon, impoundment, ditch, landfill, storage container, motor vehicle, rolling stock, or aircraft, or any site or area where a hazardous substance has been deposited, stored, disposed of, or placed, or otherwise come to be located; but does not include any consumer product in consumer use or any vessel." *Id.* § 9601(9). The courts have interpreted "facility" liberally in favor of evoking CERCLA guidelines. See, e.g., *United States v. Bliss*, 667 F. Supp. 1298, 1305 (E.D. Mo. 1987); *United States v. Conservation Chem. Co.*, 619 F. Supp. 162, 184-85 (W.D. Mo. 1985).

26. See *supra* note 15 and accompanying text.

27. CERCLA applies to all hazardous substances even if they are not hazardous "wastes." Under CERCLA, a hazardous substance means a substance that has been identified as such in other environmental statutes. 42 U.S.C. § 9601(14) (1988). Furthermore, the EPA is authorized to classify any additional substance as "hazardous" if it "may present a substantial danger to the public health or welfare of the environment." *Id.* § 9602(A). For example, asbestos has been deemed a hazardous substance for purposes of CERCLA. *United States v. Nicolet*, 712 F. Supp. 1205, 1207 (E.D. Pa. 1989).

28. *United States v. Bliss*, 667 F. Supp. 1298, 1304 (E.D. Mo. 1987); *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 573, 576 (D. Md. 1986); *United States v. Wade*, 577 F. Supp. 1326, 1333 (E.D. Pa. 1983).

responsible for creating the problem.²⁹ The four classes or categories of potentially responsible parties (PRPs) who may be liable under CERCLA include:

- (1) the current owners or operators of a hazardous waste facility;³⁰
- (2) the former owners or operators of the facility at the time of any waste disposal;
- (3) the generators of the hazardous waste; or
- (4) the transporters of the hazardous substances.³¹

Although CERCLA's list of PRPs is expansive, its definition of "owner or operator" specifically excludes any person "who, *without participating in the management* of a [hazardous waste facility], holds indicia of ownership primarily to protect his security interest."³² It appears that this exclusion from liability, known as the "security interest exception," was

29. H.R. REP. No. 1016, 96th Cong., 2d Sess., pt. 1, *reprinted in* 1980 U.S. CODE CONG. & ADMIN. NEWS 6136.

30. Courts have construed "owner and operator" under § 107(a)(1) to mean "owner or operator" because such a construction is in accord with legislative intent. See Burkhart, *supra* note 1, at 332-33. See also Guidice v. BFG Electroplating & Mfg. Co., 732 F. Supp. 566, 577-78 (W.D. Pa. 1989); Artisan Water Co. v. New Castle County, 659 F. Supp. 1269, 1280 (D. Del. 1987), *aff'd*, 851 F.2d 643 (3d Cir. 1988); *Maryland Bank & Trust Co.*, 632 F. Supp. at 577-78.

31. CERCLA § 107(a) states:

- (1) the owner and operator of a vessel or facility;
- (2) any person who at the time of disposal of any hazardous substance owned or operated any facility at which such hazardous substances were disposed of;
- (3) any person who by contract, agreement, or otherwise arranged for disposal or treatment, or arranged with a transporter for transport for disposal or treatment, of hazardous substances owned or possessed by such person, by any other party or entity, at any facility or incineration vessel owned or operated by another party or entity and containing such hazardous substances; and
- (4) any person who accepts or accepted any hazardous substances for transport to disposal or treatment facilities, incineration vessels or sites selected by such person, from which there is a release, or a threatened release which causes the incurrence of response costs, of a hazardous substance, shall be liable for —

(A) all costs of removal or remediation action incurred by the United States Government or a state or an Indian tribe not inconsistent with the national contingency plan;

(B) any other necessary costs of response incurred by any other person consistent with the national contingency plan;

(C) damages for injury to, destruction of, or loss of natural resources, including the reasonable costs of assessing such injury, destruction, or loss resulting from such a release; and

(D) the costs of any health assessment or health effects study carried out under § 9604(i) of this title.

42 U.S.C. § 9607(a) (1988).

32. *Id.* § 9601(20)(A) (emphasis added).

intended to protect lenders holding a security interest in property from being classified as CERCLA PRPs.

The primary impetus for the passage of CERCLA was to provide an effective legal mechanism for imposing liability for the clean-up of unsafe hazardous waste sites on solvent parties connected to the environmental problem.³³ During the last five years, however, secured lenders who were not responsible for the problem have found themselves ensnared in CERCLA's web of liability. This has occurred as a result of a series of judicial decisions narrowing the scope of the security interest exception. These decisions have expanded the definition of "owner or operator" of a hazardous waste site to include lending institutions in some situations.³⁴ Until recently, secured lenders that did not foreclose and did not participate in the operational decisions of a borrower were exempt from liability under CERCLA because they fell under the security interest exception.³⁵ In *United States v. Fleet Factors Corp.*,³⁶ however, the Eleventh Circuit Court of Appeals established an expanded theory of secured lender liability for a borrower's environmental problems. The court held that a secured lender is removed from the protection of the security interest exception if it is shown that the lender merely had the ability to influence a borrower's hazardous waste decision if it so chose.³⁷

This Note focuses on the interpretation and application of CERCLA's security interest exception as it applies to secured lenders. Section I discusses the overall structure of CERCLA and specifically focuses on its liability and defense provisions. Section II examines the developing liability of lenders as "owners or operators" and the applicability of the security interest exception. Section III focuses on the future of lender liability under CERCLA. Finally, Section IV outlines measures secured lenders may follow in order to minimize their risk of liability under CERCLA.

I. THE STRUCTURE OF CERCLA

A. Scope of Liability

CERCLA's primary liability provision is section 107(a).³⁸ This section allows government agencies to sue PRPs³⁹ for removal

33. Klotz & Kane, Environmental Liability Concerns for the Secured Lender 1, 1-2 (unpublished manuscript on file in the office of the Indiana Law Review).

34. See *United States v. Fleet Factors Corp.*, 901 F.2d 1550 (11th Cir. 1990); *Guidice v. BFG Electroplating & Mfg. Co.*, 732 F. Supp. 556 (W.D. Pa. 1989); *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 573 (D. Md. 1986); *United States v. Mirabile*, 15 Envtl. L. Rep. (Env. L. Inst.) 20992 (E.D. Pa. Sept. 4, 1985).

35. *Fleet Factors*, 901 F.2d at 1550.

36. *Id.*

37. *Id.* at 1558.

38. 42 U.S.C. § 9607(a). (1988). For the full text of § 107(a), see *supra* note 31.

39. 42 U.S.C. § 9607(a) (1988).

al⁴⁰ and remedial⁴¹ costs as well as damages for the destruction or loss of natural resources due to the release of hazardous substances.⁴²

The initial flurry of CERCLA litigation focused on holding PRPs liable pursuant to their status as hazardous waste "generators."⁴³ During the past five years, however, this focus has shifted. Courts have with increasing regularity imposed liability on PRPs based on their status as "owners or operators" of a hazardous waste facility.⁴⁴ In imposing such liability, courts have greatly broadened the definition of an "owner or operator," so that more classes of PRPs are subject to CERCLA's web. Courts have held lessors,⁴⁵ lessees,⁴⁶ corporate officers,⁴⁷ corporation stockholders,⁴⁸ parent corporations,⁴⁹ temporary owners and brokers,⁵⁰

40. A removal action is essentially a clean-up activity which includes studies to monitor and evaluate the release of hazardous substances. *Id.* § 9604(a).

41. A remedial action includes more long-term and permanent responses and includes confinement, ditching, trenching, recycling, segregation, and relocation of endangered residents. *Id.*

42. *Id.* §§ 9604, 9605.

43. See Comment, *CERCLA Litigation Update: The Emerging Law of Generator Liability*, 14 ENVTL. L. REP. (Envtl. L. Inst.) 10224 (June, 1984).

44. See *infra* notes 45-53 and accompanying text.

45. A passive lessor of a site is liable for response costs even if the lessor was in no way connected with the lessee's activities. *United States v. Monsanto Co.*, 858 F.2d 160 (4th Cir. 1988); *United States v. Argent Corp.*, 21 Env't Rep. Cas. (BNA) 1354 (N.D. Pa. 1984); *United States v. South Carolina Recycling & Disposal, Inc.*, 21 Env't Rep. Cas. (BNA) 1577 (D.S.C. 1984).

46. Lessees who actually cause the contamination are liable. *Caldwell v. Gurley Ref. Co.*, 755 F.2d 645 (8th Cir. 1985); *South Carolina Recycling & Disposal, Inc.*, 21 Env't Rep. Cas. (BNA) 1577 (D.S.C. 1984).

47. A corporate officer who managed the corporation's disposal activities may be personally liable for clean-up costs. *United States v. Carolawn Co.*, 21 Env't Rep. Cas. (BNA) 2124 (D.S.C. 1984); *United States v. Mottolo*, 22 Env't Rep. Cas. (BNA) 1026 (D.N.H. 1984); *United States v. Wade*, 577 F. Supp. 1326 (E.D. Pa. 1983).

48. Shareholders of a corporation are personally liable if they participate in making decisions with respect to hazardous waste disposal. *Quadriion Corp. v. Mache*, 738 F. Supp. 270 (N.D. Ill. 1990); *Kelley v. Thomas Solvent Co.*, 727 F. Supp. 1532 (W.D. Mich. 1989); *United States v. Carolawn Co.*, 21 Env't Rep. Cas. (BNA) 2124 (D.S.C. 1984).

49. In the following cases the parent corporation was liable because it had the capacity to control and manage the disposal practices of the subsidiary corporation: *State v. Bunker Hill*, 635 F. Supp. 665 (D. Idaho 1986); *Wehner v. Syntex Agribusiness, Inc.*, 616 F. Supp. 27 (E.D. Mo. 1985). See also *United States v. Stringfellow*, 661 F. Supp. 1053 (C.D. Cal. 1987) (parent corporation that merged with a wholly owned subsidiary was liable for the clean-up obligations of the subsidiary corporation). But see *Joslyn Mfg. Co. v. T. L. James & Co.*, 20 Envtl. L. Rep. (Envtl. L. Inst.) 20382 (5th Cir. Jan. 29, 1990) (disagreeing with the Second Circuit Court of Appeals and holding that parent corporations are not liable under CERCLA for the violations of their wholly owned subsidiaries).

50. *United States v. Bliss*, 667 F. Supp. 1298 (E.D. Mo. 1987) (company which acted as a broker between chemical manufacturer and disposal company was liable for response costs); *United States v. Carolawn Co.*, 21 Env't Rep. Cas. (BNA) 2124 (D.S.C. 1984) (chemical company that held title for merely one hour could be liable for clean-up costs).

present owners who did not contribute to contamination,⁵¹ lenders actively involved in a polluter's activities,⁵² and lenders who foreclosed and purchased property⁵³ to be potentially liable owners or operators under CERCLA. Courts have also held that liability under CERCLA may be applied retroactively.⁵⁴ Thus, a PRP may be liable for clean-up costs incurred prior to CERCLA's enactment.⁵⁵ In some cases, these pre-enactment costs have been substantial and have therefore imposed a great burden on the responsible party.⁵⁶

Liability under CERCLA can be enormous.⁵⁷ Responsible parties may be liable for up to \$50 million for damages to natural resources.⁵⁸ Furthermore, no ceiling on liability exists when the release or threatened release of a hazardous substance is a result of willful misconduct, willful negligence, or a refusal on the part of the disposer to provide reasonable cooperation and assistance in response activities.⁵⁹ Finally, CERCLA empowers the federal government to obtain treble damages from responsible parties who "fail without sufficient cause" to comply with clean-up plans.⁶⁰ This treble damages clause has withstood constitutional challenge⁶¹ and has been interpreted to allow treble damages when a PRP's defense is brought "in bad faith"⁶² or "without an objectively reasonable basis."⁶³

B. *The Security Interest Exception*

Although the scope of liability is encompassing, Congress sought to provide a degree of protection from liability for secured lenders through

51. *State v. Shore Realty Corp.*, 759 F.2d 1032 (2d Cir. 1985); *United States v. Tyson*, 25 Env't Rep. Cas. (BNA) 1897 (E.D. Pa. 1986); *United States v. Cauffman*, 21 Env't Rep. Cas. (BNA) 2167 (C.D. Cal. 1984).

52. *United States v. Mirabile*, 15 Envtl. L. Rep. (Envtl. L. Inst.) 20994 (E.D. Pa. Sept. 4, 1985).

53. Secured lenders who foreclose and purchase the site are liable as owners or operators under CERCLA. *Guidice v. BFG Electroplating & Mfg. Co.*, 732 F. Supp. 556 (W.D. Pa. 1989); *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 573 (D. Md. 1986).

54. *United States v. Northeastern Pharmaceutical Chem. Co.*, 810 F.2d 726, 733-35 (8th Cir. 1986); *United States v. Tyson*, 25 Env't Rep. Cas. (BNA) 1897, 1907-08 (E.D. Pa. 1986); *United States v. Shell Oil Co.*, 605 F. Supp. 1064, 1079 (D. Cal. 1985); *United States v. Ottoli & Goss, Inc.*, 630 F. Supp. 1361, 1398 (D.N.H. 1985).

55. *Northeastern Pharmaceutical Chem. Co.*, 810 F.2d at 737.

56. See *supra* note 54 and accompanying text.

57. See *supra* note 7 and accompanying text.

58. 42 U.S.C. § 9607(c)(1) (1988).

59. *Id.* § 9607(c)(2).

60. *Id.* § 9607(c)(3).

61. *Aminoil, Inc. v. United States*, 646 F. Supp. 294, 299 (C.D. Cal. 1986).

62. *Id.*

63. *Solid State Circuits, Inc. v. United States Envtl. Protection Agency*, 812 F.2d 383, 391 (8th Cir. 1987).

the "security interest exception" to CERCLA's definition of an "owner or operator." Specifically, the security interest exception provides that:

Such term [owner or operator] does not include a person, who, without participating in the management of a vessel or facility, holds indicia of ownership primarily to protect his security interest in the vessel or facility.⁶⁴

Substantial litigation has occurred over the meaning of this exclusionary provision.⁶⁵ Of critical importance to financial institutions is the question of when lenders will be included in this sanctuary. In past cases, a secured lender that did not foreclose on a borrower's property and did not participate in the debtor's daily operational decisions was held to be exempt from liability under the security interest exception.⁶⁶ Recent rulings, however, have placed these decisions in doubt.

In *United States v. Fleet Factors Corp.*⁶⁷ the Eleventh Circuit Court of Appeals held that a secured lender may be liable as a "de facto owner" of a contaminated facility if the lender participated in the financial management of the facility to an extent indicating the lender's capacity to influence the borrower's disposal decisions if it so chose.⁶⁸ This decision significantly narrowed the scope of the security interest exception so that the extent of the future application of this provision is now unclear.

C. Standard of Liability Under CERCLA

Although CERCLA did not specifically indicate the standard of liability imposed upon responsible parties, courts have consistently interpreted the Act to provide that responsible parties are subject to a strict liability standard.⁶⁹ Courts have reasoned that such an interpretation

64. 42 U.S.C. § 9601(20)(A) (1988).

65. See generally *United States v. Fleet Factors*, 901 F.2d 1550 (11th Cir. 1990); *Guidice v. BFG Electroplating & Mfg. Co.*, 732 F. Supp. 556 (W.D. Pa. 1989); *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 573 (D. Md. 1986); *In re T.P. Long Chem., Inc.*, 45 Bankr. 278 (N.D. Ohio 1985); *United States v. Mirabile*, 15 Envtl. L. Rep. (Envtl. L. Inst.) 20992 (E.D. Pa. Sept. 4, 1985).

66. *Guidice*, 732 F. Supp. at 556; *United States v. New Castle County*, 727 F. Supp. 854 (D. Del. 1989); *Rockwell Int'l v. I.U. Int'l Corp.*, 702 F. Supp. 1384 (N.D. Ill. 1988).

67. 901 F.2d 1550 (11th Cir. 1990); *Maryland Bank & Trust Co.*, 632 F. Supp. at 573.

68. *Fleet Factors*, 901 F.2d at 1558.

69. E.g., *United States v. Monsanto Co.*, 858 F.2d 160, 167 (4th Cir. 1988); *United States v. Northeastern Pharmaceutical Chem. Co.*, 810 F.2d 726, 743 (8th Cir. 1986); *Kelley v. Thomas Solvent Co.*, 727 F. Supp. 1532, 1541 (W.D. Mich. 1989); *Maryland Bank & Trust Co.*, 632 F. Supp. at 576; *United States v. Dickerson*, 640 F. Supp. 448, 451 (D. Md. 1986); *United States v. Ward*, 618 F. Supp. 884, 893 (E.D.N.C. 1985); *City of Philadelphia v. Stepan Chem. Co.*, 544 F. Supp. 1135, 1148 (E.D. Pa. 1982).

is consistent with the Act's legislative intent.⁷⁰ As a result, a party may be liable under CERCLA even though its actions were neither intentional nor negligent.

Several courts have also interpreted CERCLA to impose liability without proof of causation between a hazardous waste generator's conduct and the release or threatened release of hazardous waste.⁷¹ In *United States v. Wade*,⁷² the court stated:

[T]o require a plaintiff under CERCLA to "fingerprint" wastes is to eviscerate the statute . . . The only required nexus between the defendant and the site is that the defendant has dumped his waste there and the [type of] hazardous substances found in the defendant's waste are also found at the site.⁷³

Consequently, unless the defendant can prove that he falls under one of CERCLA's narrow defenses,⁷⁴ the defendant will be liable if the disposed waste is of the type involved in the pollution of the site.⁷⁵

D. Apportionment of Liability

CERCLA does not mention whether responsible parties are subject to joint and several liability. Although early versions of the legislation provided that defendants would be jointly and severally liable for clean-up costs, the final version of the bill deleted that provision.⁷⁶ Courts nonetheless, have held that PRPs may be held jointly and severally liable under CERCLA.⁷⁷ Courts have reasoned that this interpretation comports with CERCLA's intended purpose because it allows the government better flexibility in recovering hazardous waste response costs.⁷⁸ Con-

70. *United States v. Price*, 577 F. Supp. 1103, 1114 (D.N.J. 1983).

71. *Monsanto*, 858 F.2d at 169-70; *State v. Shore Realty Corp.*, 759 F.2d 1032, 1044-45 (2d Cir. 1985); *United States v. Ottali & Goss, Inc.*, 630 F. Supp. 1361, 1402 (D.N.H. 1985); *United States v. Conservation Chem. Co.*, 619 F. Supp. 162, 236 (W.D. Mo. 1985); *United States v. Wade*, 577 F. Supp. 1326 (E.D. Pa. 1983).

72. 577 F. Supp. 1326 (E.D. Pa. 1983).

73. *Id.* at 1332-33.

74. For a discussion of the affirmative defenses available under CERCLA, see *infra* notes 83-86 and accompanying text.

75. See Klotz & Kane, *supra* note 33, at 9.

76. *United States v. Chem-Dyne Corp.*, 572 F. Supp. 802, 807 (S.D. Ohio 1983).

77. *United States v. Monsanto Corp.*, 858 F.2d 160, 167 (4th Cir. 1988); *Kelley v. Thomas Solvent Co.*, 727 F. Supp. 1532, 1552 (W.D. Mich. 1989); *United States v. Stringfellow*, 661 F. Supp. 1053, 1061 (C.D. Cal. 1987); *State v. Bunker Hill Co.*, 635 F. Supp. 665, 676 (D. Idaho 1986); *United States v. A & F Materials*, 578 F. Supp. 1249, 1254-55 (S.D. Ill. 1984); *Chem-Dyne Corp.*, 572 F. Supp. at 807-08.

78. *A & F Materials*, 578 F. Supp. at 1253-55; *Chem-Dyne Corp.*, 572 F. Supp. at 807-08.

sequently, courts will impose joint and several liability on defendants unless the harm is shown to be divisible.⁷⁹

As originally enacted, CERCLA was also silent concerning whether responsible parties have a right to contribution. Courts nonetheless determined that based upon common-law principles, a party who has incurred disproportionate liability under CERCLA has a right to contribution against other responsible parties.⁸⁰ In 1986, SARA codified this right to contribution.⁸¹ At least one court has held that a PRP may sue for contribution for necessary clean-up costs *before* the federal or state government decides to commence an action against a responsible parties.⁸²

E. Affirmative Defenses Under CERCLA

Three meager defenses are available to persons falling under CERCLA's broad liability scheme. Section 107(b) exculpates a person otherwise liable under CERCLA who can demonstrate that the release or threatened release of hazardous substances was caused solely by: (1) an act of God; (2) an act of war; or (3) an act of a third party having no contractual relation to the person.⁸³ Courts have construed these

79. *Monsanto*, 858 F.2d at 167; *Kelley*, 727 F. Supp. at 1552; *Stringfellow*, 661 F. Supp. at 1061.

80. *Caldwell v. Gurley Ref. Co.*, 755 F.2d 645, 651-52 (8th Cir. 1985); *State v. ASARCO, Inc.*, 608 F. Supp. 1484, 1491 (D. Col. 1985).

81. 42 U.S.C. § 9607(e)(2) (1988).

82. *Versatile Metals, Inc. v. Union Corp.*, 693 F. Supp. 1563, 1571 (E.D. Pa. 1988).

83. 42 U.S.C. § 9607(b) states:

There shall be no liability under subsection (a) of this section for a person otherwise liable who can establish by a preponderance of the evidence that the release or threat of release of a hazardous substance and the damages resulting therefrom were caused solely by—

- (1) an act of God;
- (2) an act of war;

(3) an act or omission of a third-party other than an employee or agent of the defendant, or than one whose act or omission occurs in connection with a contractual relationship, existing directly or indirectly, with a defendant (except where the sole contractual arrangement arises from a published tariff and acceptance for carriage by a common carrier by rail), if the defendant establishes by a preponderance of the evidence that (a) he exercised due care with respect to the hazardous substance concerned, taking into consideration the characteristics of such hazardous substance, in light of all relevant facts and circumstances, and (b) he took precautions against foreseeable acts or omissions of any such third party and the consequences that could foreseeably result from such acts or omissions; or

- (4) any combination of the foregoing paragraphs.

defenses narrowly and have seldom allowed parties to successfully invoke these provisions.⁸⁴

The third party defense is the most frequently asserted affirmative defense. In 1986, SARA slightly broadened this defense by providing a detailed definition of what constitutes a "contractual relationship" for purposes of section 107(b)(3).⁸⁵ This amendment essentially provides that before a landowner can utilize the third party defense, he must be able to demonstrate that he made all reasonable inquiries into the previous activities on the property and had "no reason to know" that any hazardous substances were disposed of at the site.⁸⁶ In reality, this "innocent landowner defense" is also extremely limited in its scope because the defendant must show that he had neither actual nor constructive knowledge of any hazardous waste threat at the time the property was acquired.

II. THE DEVELOPMENT OF LENDER LIABILITY UNDER CERCLA

When CERCLA was enacted, the security interest exception⁸⁷ seemed to afford lenders a safe harbor from being ensnared in the Act's web of liability. This provision excludes from liability a secured creditor who "without participating in the management" of the facility "holds indicia of ownership primarily to protect his security interests."⁸⁸ Judicial in-

84. *United States v. Monsanto Corp.*, 858 F.2d 160, 168-69 (4th Cir. 1988); *United States v. Stringfellow*, 661 F. Supp. 1053, 1061 (C.D. Cal. 1989); *United States v. Argent Corp.*, 21 Env't Rep. Cas. (BNA) 1354, 1356 (N.D. Pa. 1984).

85. 42 U.S.C. § 9601(35)(A) (1988) states:

The term "contractual relationship," for the purpose of section 9607(b)(3) of this title, includes, but is not limited to, land contracts, deeds or other instruments transferring title or possession, unless the real property on which the facility concerned is located was acquired by the defendant after the disposal or placement of the hazardous substance on, in, or at the facility, and one or more of the circumstances described in clause (i), (ii), or (iii) is also established by the defendant by a preponderance of the evidence:

(i) At the time the defendant acquired the facility the defendant did not know and had no reason to know that any hazardous substance which is the subject of the release or threatened release was disposed of on, in, or at the facility.

(ii) The defendant is a government entity which acquired the facility by escheat, or through any other involuntary transfer or acquisition, or through the exercise of imminent domain authority by purchase or condemnation.

(iii) The defendant acquired the facility by inheritance or bequest.

In addition to the foregoing, the defendant must establish that he has satisfied the requirements of § 9607(b)(3)(a) and (b) of this title.

86. 42 U.S.C. § 9601(35)(A)(i) (1988).

87. *Id.* § 9601(20)(A).

88. *Id.*

terpretation of the secured interest exception, however, has had a checkered history and result. Courts disagree as to when and if this exception applies to a foreclosing lender who acquires ownership of the property to satisfy a secured debt.⁸⁹ In addition, courts have established conflicting standards of what constitutes impermissible participation in a debtor's management to nullify the security interest exception.⁹⁰ Unfortunately, Congress has provided no guidance in resolving these issues, and the CERCLA and SARA provisions do not address these interpretation concerns.

A. *United States v. Mirabile*

*United States v. Mirabile*⁹¹ was the first case to specifically interpret the scope of the security interest exception. In *Mirabile*, the United States brought suit against the Mirabiles, the current owners and operators of a site containing hazardous waste, to recover clean-up costs associated with the removal of hazardous substances. The Mirabiles then joined, among others, American Bank & Trust Company (ABT) and Mellon Bank National Association (Mellon Bank). Both of these banks made secured loans to the site's former owners.

In 1973, ABT loaned money to Mangels Industries, a paint manufacturing company. The note was secured in part by a mortgage on the property. In 1976, Turco Coatings, Inc. acquired Mangels Industries and borrowed money from Girard Bank, the predecessor in interest of Mellon Bank. This loan was secured by Turco's assets and inventory. Turco generated and deposited hazardous wastes at the site until 1980 when it ceased its operations due to financial difficulties.

In 1981, ABT foreclosed on the Turco property. It was then the highest bidder at the sheriff's sale and subsequently notified the sheriff that it intended to take title to the property. In December, 1981, however, ABT assigned its bid to the Mirabiles who accepted title to the property.

When sued by the EPA, the Mirabiles joined ABT and Mellon Bank as third party defendants, claiming that the banks were also PRPs under

89. See *In re Bergsoe Metal Corp.*, 910 F.2d 668 (9th Cir. 1990); *United States v. Fleet Factors Corp.*, 901 F.2d 1550 (11th Cir. 1990); *Guidice v. BFG Electroplating Mfg. Co.*, 732 F. Supp. 556 (W.D. Pa. 1989); *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 573 (D. Md. 1986); *United States v. Mirabile*, 15 Envtl. L. Rep. (Envtl. L. Inst.) 20992 (E.D. Pa. Sept. 4, 1985); *In re T. P. Long Chem., Inc.*, 45 Bankr. 278 (N.D. Ohio 1985).

90. See *supra* note 89.

91. 15 Envtl. L. Rep. (Envtl. L. Inst.) 20992 (E.D. Pa. Sept. 4, 1985).

CERCLA as a result of their activities concerning the Turco Site. ABT and Mellon Bank then counterclaimed against the United States based upon the involvement of the Small Business Administration (SBA) in the site.⁹² ABT, Mellon Bank, and the SBA moved for summary judgment claiming nonliability based upon the security interest exception.

1. ABT's Motion for Summary Judgment.—ABT moved for summary judgment on two grounds. First, it contended that under Pennsylvania law, its successful bid at the sheriff's sale gave it only equitable title to the property which never evolved into legal title because it subsequently assigned its bid to the Mirabiles. Second, ABT asserted that its activities at the site were undertaken merely to protect its security interest in the Turco Site and that it never "participated in the management" of that site.⁹³

The court agreed with ABT's first argument and held that the passage of either legal or equitable title was irrelevant to the applicability of the security interest exception. The court stated that:

[It] need not resolve the issue of whether, under Pennsylvania law, ABT's successful bid at the Sheriff's sale technically vested ABT with ownership as defined by the statute. Regardless of the nature of the title received by ABT, its actions with respect to the foreclosure were plainly undertaken in an effort to protect its security interest in the property.⁹⁴

Therefore, the court concluded that ABT's foreclosure did not nullify its status under the security interest exception.

The court also agreed with ABT's second contention and held that the limited actions taken by ABT after foreclosure did not constitute "participation in the management of the site."⁹⁵ The court opined that, "it would appear that before a secured creditor such as ABT may be held liable, it must, at a minimum, participate in the day-to-day operational aspects of the site."⁹⁶ ABT's only actions with reference to the Turco Site were securing it against vandalism, inquiring as to clean-up costs, and showing the site to prospective buyers. The court found that these actions were only routine steps designed to prevent further depreciation of the property and could not be deemed participation in

92. In July of 1979, the SBA loaned \$150,000 to Turco which was to be applied to specific debts. The loan was secured by a second lien security interest in Turco machinery, equipment, inventory, accounts receivable, and property.

93. *United States v. Mirabile*, 15 Envtl. L. Rep. (Envtl. L. Inst.) 20992, 20996 (E.D. Pa. Sept. 4, 1985).

94. *Id.*

95. *Id.*

96. *Id.*

the day-to-day activities of the site.⁹⁷ Consequently, ABT's motion for summary judgment was granted, and ABT was not deemed a PRP.⁹⁸

2. *SBA's Motion for Summary Judgment.*—Although the SBA took neither legal nor equitable title to the Turco Site, the loan agreement allowed the SBA to participate in some of the financial management decisions of Turco. The Mirabiles contended that because the SBA was empowered to place restrictions on loan proceeds, the SBA may have prevented Turco from properly disposing of the hazardous wastes on the site.

The court rejected the Mirabiles' argument on two grounds. First, the court noted that there was no evidence that the SBA actually chose to assert its financial power over Turco.⁹⁹ Second, the court noted that participation in purely *financial* aspects of an operation is not sufficient to bring a lender within the scope of CERCLA liability.¹⁰⁰ Because the SBA's influence was limited to financial matters, the court granted its motion for summary judgment.¹⁰¹

3. *Mellon Bank's Motion for Summary Judgment.*—In contrast to the ABT and SBA motions, the court found itself faced with a "cloudier situation" when it considered Mellon Bank's motion for summary judgment.¹⁰² Mellon Bank, through its predecessor in interest, Girard Bank, held a security interest in Turco's inventory and assets. After Turco defaulted on its loan, Girard became involved in Turco's operations. Initially, it placed one of its loan officers on an advisory board established to oversee Turco's operations. Later, the Bank became more involved by making weekly visits and instructing the company on manufacturing, personnel, and sales matters.

The Mirabiles asserted that Mellon Bank's involvement with Turco brought it within the scope of CERCLA liability. The court agreed with the Mirabiles and held that although "[t]he reed upon which the Mirabiles seek to impose liability . . . is slender indeed,"¹⁰³ a genuine issue of fact was presented as to whether Mellon Bank, through its predecessor Girard Bank, engaged in the sort of *participation in management* which would bring a secured creditor within the scope of CERCLA liability.¹⁰⁴ Therefore, the court denied Mellon Bank's motion for summary judgment.¹⁰⁵

97. *Id.*

98. *Id.*

99. *Id.*

100. *Id.*

101. *Id.*

102. *Id.* at 20997.

103. *Id.*

104. *Id.*

105. *Id.*

The court's decision was a bittersweet result for the lending industry. On one hand, the court held that mere participation in the *financial* aspects of a borrower's waste disposal practices is not sufficient for the imposition of CERCLA liability.¹⁰⁶ Additionally, the court noted that a lender's foreclosure and subsequent acquisition of property will not necessarily remove it from the safe harbor afforded by the security interest exception.¹⁰⁷ On the other hand, lenders must be aware that they may incur liability under CERCLA if they become actively involved in a borrower's operations in order to protect a secured interest.

B. *United States v. Maryland Bank & Trust Company*

Shortly after *Mirabile*, the issue of lender liability was again addressed in *United States v. Maryland Bank & Trust Co.*¹⁰⁸ Contrary to *Mirabile*, the court in *Maryland Bank* held that the security interest exception does not protect a mortgagee who forecloses and subsequently takes title to the property.¹⁰⁹

In 1980, Maryland Bank & Trust Company (MBT) made a secured loan to Mark McCleod, the owner and operator of the California Maryland Drum (CMD) site. In 1981, after McCleod failed to make payments, MBT foreclosed on the CMD site and purchased the property at the foreclosure sale. In 1983, the EPA found that hazardous wastes were improperly disposed of at the CMD site. After the MBT refused the EPA's order to clean up the site, the EPA proceeded to clean the site itself at a cost of over \$550,000. The EPA then sued MBT for payment claiming that it was liable under CERCLA as the present owner and operator of the facility.

MBT's primary defense was that it was not liable to the EPA because it was protected by CERCLA's security interest exception.¹¹⁰ In its summary judgment motion, MBT asserted that it was not liable as a present owner or operator of the CMD site because it had foreclosed and took title to the property merely to protect its security interests. The United States contended, however, that MBT was liable as a present owner and operator of the site, and it was not entitled to any exemption because it no longer held a security interest in the property.

The court agreed with the Government's position and held that MBT was liable as a responsible party under CERCLA.¹¹¹ The court supported

106. *Id.*

107. *Id.*

108. 632 F. Supp. 573 (D. Md. 1986).

109. *Id.* at 579-80.

110. *Id.* at 579.

111. *Id.*

its finding by utilizing a literal reading of the security interest exception. It stated that the “verb tense of the security interest exception is critical. The security interest must exist *at the time* of the clean-up” in order to afford a lender any protection.¹¹² The court consequently ruled that the security interest exception terminates upon a lender’s foreclosure and subsequent purchase of the property because the security interest is destroyed as the lender’s interest ripens into full title.¹¹³

Next, the court asserted that policy considerations played a role in its determination of liability. The court reasoned that:

Under the scenario put forward by the bank, the federal government would alone shoulder the cost of cleaning up the site, while the former mortgagee-turned-owner, would benefit from the clean-up by the increased value of the now unpolluted land. At the foreclosure, the mortgagee could acquire the property cheaply. All other prospective purchasers would be faced with potential CERCLA liability, and would shy away from the sale. Yet once the property has been cleared at the taxpayers’ expense and becomes marketable, the mortgagee-turned-owner would be in a position to sell the site at a profit.

In essence, the [bank’s] position would convert CERCLA into an insurance scheme for financial institutions, protecting them against possible losses due to the security of loans with polluted properties.¹¹⁴

The court further noted that lenders were not without protection under such a liability scheme. It stated that banks could protect themselves by conducting an environmental audit before lending any money. Based upon these findings, the court granted the Government’s motion for summary judgment pertaining to MBT’s liability under section 107(a)(1).¹¹⁵

C. *Guidice v. BFG Electroplating & Manufacturing Company*

In *Guidice v. BFG Electroplating & Manufacturing Co.*,¹¹⁶ the District Court for the Western District of Pennsylvania was faced with essentially the same interpretation issues involving the security interest exception that were examined in *Mirabile*. Contrary to the *Mirabile* decision, however, *Guidice* interpreted the security interest exception somewhat narrowly.

112. *Id.* (emphasis in original).

113. *Id.*

114. *Id.* at 580.

115. *Id.* at 582.

116. 732 F. Supp. 556 (W.D. Pa. 1989).

In *Guidice*, the National Bank of the Commonwealth issued a secured loan to Berlin Metal Polishers (Berlin Metal) in 1975. After Berlin Metal defaulted on its obligation to the Bank in 1980, the Bank sent representatives to tour the Berlin Metal plant and to meet with officials to discuss management. The Bank was informed of the number of work shifts, the status of Berlin Metal's accounts, the composition of management, and the presence of raw materials. At the meeting, the Bank suggested that the company take out a loan guaranteed by the SBA to pay off the monies that Berlin Metal owed the Bank. The Bank subsequently submitted the SBA loan application on behalf of Berlin Metal and recommended its approval, but the loan was refused.

A few months later, a Bank officer was contacted by a prospective purchaser of the Berlin Metal site. The Bank officer independently discussed the matter with city officials. The sale was never made, however, and, in 1981, Berlin ceased operations. At this time, a Bank officer was again sent to Berlin Metal to discuss restructuring the loan. Finally, in June, 1981, the Bank foreclosed on the property. The Bank subsequently bought the Berlin Metal facility at the sheriff's sale in 1982 and sold the property eight months later. In 1986, local residents sued BFG (the present property owner) and alleged that contamination from the Berlin Metal site was causing personal injuries. BFG then filed a third party complaint against the Bank alleging that the Bank was liable for response costs as a former "owner or operator" of the Berlin Metal site.

It was necessary for the court to address two separate arguments in order to render a decision on the Bank's motion for summary judgment. First, BFG contended that the Bank was liable as an owner or operator *before* it foreclosed on the Berlin Metal property. The Bank responded that it was immune from liability, however, because it fell under the security interest exception. At issue was whether the Bank's "participation in the management" of Berlin Metal was sufficient to remove it from the security interest exception.

The court chose to espouse the narrow interpretation of "participating in management" previously outlined in *Mirabile*. The court held that a secured creditor may provide financial assistance and may even provide isolated instances of specific management advice to its debtors without risking CERCLA liability.¹¹⁷ Applying this standard, the court found that the Bank's activities prior to foreclosure were insufficient to nullify the security interest exception.¹¹⁸ The court reasoned that none of the

117. *Id.* at 561. See also *United States v. Mirabile*, 15 Envtl. L. Rep. (Envtl. L. Inst.) 20992 (E.D. Pa. Sept. 4, 1985).

118. *Guidice v. BFG Electroplating & Mfg. Co.*, 732 F. Supp. 556, 562 (W.D. Pa. 1989).

Bank's activities concerning the Berlin property suggested that the Bank "controlled operational, production, or waste disposal activities" of the facility.¹¹⁹

Addressing the second issue, the court found that although the Bank could claim protection under the security interest exception with respect to its actions *before* foreclosure, the Bank could not claim protection under the security interest exception with respect to its activities *after* it foreclosed on the Berlin property.¹²⁰ Recognizing a divergence in case law, the court chose to adopt the position expressed in *United States v. Maryland Bank & Trust Co.*, rather than that established in *United States v. Mirabile*. Consequently, the court held that "when a lender is the successful purchaser at a foreclosure sale, the lender should be liable to the same extent as any other bidder at the sale would have been."¹²¹

The court supported its conclusion by pointing to the amendments to CERCLA in 1986. The court found that while Congress excluded other entities from CERCLA liability through the amendments, it did not "simultaneously amend the statute to exclude from liability lenders who acquired property through foreclosure."¹²² The court opined that this indicated a congressional intent to hold those lenders liable as owners.¹²³ In applying its interpretation to the case at hand, the court found that because the Bank purchased the Berlin property at the foreclosure sale, it was removed from the protection of the security interest exception.¹²⁴ Therefore, the court held that the third party defendant could proceed with its cause of action against the Bank for the period that the Bank had record title of the property.¹²⁵

D. *United States v. Fleet Factors Corporation*

*United States v. Fleet Factors Corp.*¹²⁶ is the most recent decision specifically construing the security interest exception's participation in management clause. In *Fleet Factors*, the Eleventh Circuit Court of Appeals held that a secured lender may incur CERCLA liability as a *de facto* owner if the lender participates in the borrower's financial management to an extent that indicates an ability to influence the borrower's hazardous waste disposal decisions.¹²⁷

119. *Id.*

120. *Id.* at 563.

121. *Id.*

122. *Id.*

123. *Id.*

124. *Id.*

125. *Id.* at 564.

126. 901 F.2d 1550 (11th Cir. 1990).

127. *Id.* at 1558.

From 1963 until 1981, Swainsboro Print Works, Inc. (SPW) or its predecessor in interest operated a cloth printing facility. In 1976, SPW and the Fleet Factors Corporation (Fleet Factors) entered into a factoring agreement in which Fleet Factors agreed to advance funds against the assignment of SPW's accounts receivable. As collateral for these advances, Fleet Factors obtained a security interest in all of SPW's equipment, inventory, and fixtures. In addition, Fleet Factors was granted a security interest in the SPW plant and property. In early 1981, Fleet Factors ceased advancing funds to SPW because SPW's debt to Fleet Factors exceeded the value of its accounts receivable. On February 27, 1981, SPW ceased operations and began to liquidate its inventory. By December 1981, SPW was adjudicated a bankrupt under Chapter 7, and a trustee assumed title and control of the facility.

After obtaining court approval in 1982, Fleet Factors foreclosed on its security interest in some of SPW's inventory and equipment¹²⁸ and contracted with Baldwin Industrial Liquidators (Baldwin) to conduct an auction of the collateral. Fleet Factors also contracted with Nix Riggers (Nix) to remove the unsold equipment from the SPW facility. Nix left the facility by the end of 1983.

Upon discovering that the SPW site was contaminated with hazardous substances, in 1984, the EPA cleaned up the facility at a cost of nearly \$400,000. In 1987, the facility was conveyed to Emanuel County, Georgia at a foreclosure sale resulting from SPW's failure to pay taxes. The United States subsequently sued Fleet Factors, among others, to recover the response costs incurred in cleaning up the SPW facility.

On appeal, the Eleventh Circuit addressed the issue of whether Fleet Factors's participation in the management of the SPW site caused it to incur liability as a past owner or operator of the site under CERCLA section 107(a)(2).¹²⁹ It was undisputed that Fleet Factors had a security interest in the SPW facility. What was in dispute, however, was whether Fleet Factors was exempt from liability because of its status under the security interest exception. The Government argued that Fleet Factors's activities at SPW nullified its status under the security interest exception

128. Fleet Factors never foreclosed on its security interest in the SPW plant and property, however.

129. The district court, in ruling on the Government's and Fleet Factors's Cross Motions for summary judgment, held that: (1) Fleet Factors's activities *before* February 27, 1981 at the SPW facility did not rise to the level of "participation in management" sufficient to incur liability as a past owner or operator under CERCLA § 107(a)(2), and (2) a genuine dispute existed as to whether Fleet Factors's activities *after* February 27, 1981 at the facility rose to a level of "participation in management" sufficient to incur liability as a past owner or operator under CERCLA. *United States v. Fleet Factors Corp.*, 724 F. Supp. 955, 960-61 (S.D. Ga. 1988), *aff'd*, 901 F.2d 1550 (11th Cir. 1990).

because Fleet Factors "participated in the management" of the facility.

The Government urged the court to adopt a narrow and strict interpretation of the security interest exception "that excludes from its protection any secured creditor that participates in *any* manner in the management of the facility."¹³⁰ The court quickly rejected this argument and stated that such an interpretation would "largely eviscerate the exemption Congress intended to afford secured creditors."¹³¹ Fleet Factors in turn suggested that the court adopt an interpretation similar to the one espoused in *United States v. Mirabile* which delineated between permissible participation in the financial management of a facility and impermissible management in the day-to-day operational management of a facility.¹³²

The court of appeals chose to reject both of the proposed definitions and instead, delineated one of its own. The court of appeals found the *Mirabile* construction of the security interest exception "too permissive" toward secured creditors.¹³³ The court reasoned that the *Mirabile* court's broad interpretation would essentially require a secured creditor to be involved in the daily operations of a facility before it would incur any liability.¹³⁴ The court opined that such a construction would essentially render the security interest exception meaningless because persons whose activities rose to such a level were already liable as "operators" under section 9607(a)(2).¹³⁵ The court of appeals advocated the following standard:

[A] secured creditor may incur Section 9607(a)(2) liability, without being an operator, by participating in the *financial* [or operational] management of a facility to a degree indicating a capacity to influence the [borrower's] treatment of hazardous wastes. It is not necessary for the secured creditor actually to involve itself in the day-to-day operations of the facility in order to be liable . . . [n]or is it necessary for the secured creditor to participate in management decisions related to hazardous waste. Rather, a

130. *United States v. Fleet Factors, Corp.*, 901 F.2d 1550, 1556 (11th Cir. 1990).

131. *Id.*

132. The district court in *Fleet Factors* chose to adopt such an interpretation by relying on the holding in *Mirabile*. The district court interpreted the statutory language to permit secured creditors to "provide financial assistance in general, and even isolated instances of specific, management advice to its debtors without risking CERCLA liability if the secured creditor does not participate in the day-to-day management of the business or facility either before or after the business ceases operations." *United States v. Fleet Factors, Corp.*, 724 F. Supp. 955, 960 (S.D. Ga. 1988), *aff'd*, 901 F.2d 1550 (11th Cir. 1990).

133. *United States v. Fleet Factors, Inc.*, 901 F.2d 1550, 1557 (11th Cir. 1990).

134. *Id.*

135. *Id.*

secured creditor will be liable if its involvement with the management of the facility is sufficiently broad to support the inference that it could affect hazardous waste disposal decisions if it so chose.¹³⁶

The court listed several policy interests furthered by its interpretation of the security interest exception. First, it stated that lenders could still monitor any aspect of a debtor's business and become occasionally involved in financial decisions relating to the protection of their security interests without incurring liability.¹³⁷ Second, the court suggested that its ruling would encourage potential creditors to investigate thoroughly the waste treatment operations of potential debtors before making any loans and would provide powerful incentives for potential debtors to improve their handling of hazardous waste.¹³⁸ Finally, the court asserted that awareness of potential CERCLA liability will encourage creditors to continually monitor the existing operations of debtors and insist upon compliance with acceptable disposal procedures.¹³⁹ The court stated that society will benefit from such a result.¹⁴⁰

The court next applied its interpretation of the security interest exception to the factual situation at hand. First, the court agreed with the court below and found that Fleet Factors's involvement with the facility before February 27, 1981 was within the bounds of the security interest exception.¹⁴¹ The court also held, however, that Fleet Factors's alleged activities after February 27, 1981, if proven, would be sufficient to remove it from the protection of the security interest exception.¹⁴² The Government alleged that after this date, Fleet Factors required SPW to seek its approval before shipping inventory, established prices for excess inventory, stipulated when and to whom inventory would be shipped, determined when employees would be laid off, supervised administrative activities, received and processed SPW tax forms, controlled access to the facility, and contracted with Baldwin to dispose of SPW's fixtures and equipment. The court found that these activities removed Fleet Factors from the security interest exception.¹⁴³ After its lengthy analysis, the court denied Fleet Factors's summary judgment motion as

136. *Id.* at 1557-58 (emphasis added).

137. *Id.* at 1558.

138. *Id.*

139. *Id.*

140. *Id.*

141. SPW ceased operations and began to liquidate its inventory on February 27, 1981.

142. *United States v. Fleet Factors, Corp.*, 901 F.2d 1550, 1559 (11th Cir. 1990).

143. *Id.*

to its liability as a past owner or operator under CERCLA.¹⁴⁴ The court then remanded the case for further proceedings because disputed issues of material fact remained which had to be resolved and applied to the new interpretation of the security interest exception.¹⁴⁵

E. In re Bergsoe Metal Corporation

Only three months after the *Fleet Factors* decision, another United States Court of Appeals had the opportunity to interpret the "participating in management" clause of CERCLA's security interest exception. In *In re Bergsoe Metal Corp.*,¹⁴⁶ the Ninth Circuit Court of Appeals implicitly rejected the interpretation of the security interest exception outlined in *Fleet Factors*. The court side-stepped the opportunity to delineate its own interpretation of the clause, however.

In 1978, Bergsoe Metals, a lead recycling corporation, contacted the Port of St. Helens (Port), a municipal corporation, to discuss the building of a lead recycling facility in St. Helens. In 1979, the Port agreed to issue bonds and sell Bergsoe fifty acres of land on which to construct the plant. In exchange, the Port received a promissory note for \$400,000 and a mortgage on the property.

Through a series of interlocking transactions after the initial agreement, Bergsoe, the Port, and the United States National Bank of Oregon completed the arrangement. At the center of this financing scheme were the revenue bonds issued by the Port. The Bank held these bonds in trust for the bondholders, and Bergsoe was required to pay the money owed on the bonds to the Bank rather than the Port. Although the Port was technically the owner of the property, the Bank held the deed in escrow so that Bergsoe had the right to repurchase the property for \$100 once the bonds were paid in full.

Soon after the Bergsoe plant began its operations in 1982, it began experiencing financial difficulties. In September of 1983, the Bank declared Bergsoe to be in default. Subsequently, the Bank and Bergsoe agreed to a "workout arrangement" in which Front Street Management Corporation would manage the facility. After the plant fared no better under new management, it was forced into bankruptcy in 1986. By that time, hazardous substances were discovered to have contaminated the facility.

In 1987, the Bank filed suit against East Asiatic Company (EAC), the company that owned Bergsoe, to collect on its debts. The Bank also sought a declaratory judgment that EAC was liable for the costs of

144. *Id.* at 1560.

145. *Id.*

146. 910 F.2d 668 (9th Cir. 1990).

cleaning up the plant. Bergsoe and EAC counterclaimed and brought third party actions against the Bank and the Port, arguing that they were liable for the clean-up costs under CERCLA. The Port moved for summary judgment, claiming that it was not liable for any response costs because it fell under the protection of CERCLA's security interest exception. The bankruptcy court granted the Port's motion and the district court affirmed. The EAC then appealed.

On appeal, the Ninth Circuit Court of Appeals first ruled that even though the Port was "technically" the owner of the Bergsoe property, it could claim protection under the security interest exception because it acted as a secured creditor rather than as an owner.¹⁴⁷ In so ruling, the court ignored previous decisions that strictly interpreted CERCLA and imposed liability upon parties who were merely temporary owners or brokers of a polluted property.¹⁴⁸ The court supported its finding by noting that the Bank "essentially financed" the Bergsoe plant while the Port's "only involvement was to give its approval to the project and to issue the bonds that served as the vehicle for the financing."¹⁴⁹ It further noted that the Port only received the warranty deed as part of a transaction in which the sole purpose was to provide financing for the plant.¹⁵⁰

Next, the court addressed whether the Port sufficiently participated in the management of the Bergsoe facility to remove it from the security interest exception. Although the court had the opportunity to define the "participating in management" phrase, it chose not to do so because it was clear that the Port fell within the boundaries of the security interest exception. The court stated:

We leave for another day the establishment of a Ninth Circuit rule on this difficult issue. It is clear from the statute that, whatever the precise parameters of "participation," there must be *some* actual management of the facility before a secured creditor will fall outside the exception. Here there was none, and we therefore need not engage in line drawing.¹⁵¹

The court supported its conclusion by pointing out that there was no evidence that the Port participated in any of Bergsoe's management.¹⁵²

147. *Id.* at 671.

148. *United States v. Bliss*, 667 F. Supp. 1298 (E.D. Mo. 1989) (company which acted as a broker between chemical manufacturer and disposal company was liable for response costs); *United States v. Carolawn Co.*, 21 Env't Rep. Cas. (BNA) 2124 (D.S.C. 1984) (chemical company that held title for merely one hour could be liable for clean-up costs).

149. *In re Bergsoe Metal Corp.*, 910 F.2d 668, 671 (9th Cir. 1990).

150. *Id.*

151. *Id.* at 672.

152. *Id.* at 673.

The court first asserted that the Port's involvement in the deal's "negotiations" did not constitute management participation.¹⁵³ The court reasoned that to hold the Port liable for its involvement in the negotiations would render the security interest exception meaningless because secured creditors always have some input on the planning stages of a project.¹⁵⁴ Similarly, the court concluded that the Port's right to inspect and re-enter the Bergsoe property could not be classified as management participation.¹⁵⁵ The court opined that "what is critical is not what rights the Port had, but what it did."¹⁵⁶ Because the Port never chose to exercise its rights, it could not be deemed to have actively participated in Bergsoe's management decisions.¹⁵⁷ Finally, the court noted that no evidence existed that the Port participated in the decision to hire Front Street Management Corporation to manage the facility.¹⁵⁸ Those negotiations were entirely between the Bank, Bergsoe, and Front Street. Finding that the Port had no actual participation in the Bergsoe facility, the court granted the Port's motion for summary judgment and held that it was not liable for clean-up costs as an owner or operator under CERCLA.¹⁵⁹

III. THE FUTURE OF LENDER LIABILITY

A. *Lenders Who Foreclose And Acquire Title*

It is clear that the *Mirabile* court on the one hand and the *Maryland Bank* and *Guidice* courts on the other, approached the issue of lender liability for foreclosing lenders from opposite directions. In *Mirabile*, the court held that if a lender did not actively participate in the borrower's operations, it would not be liable for clean-up costs even if it foreclosed and acquired title.¹⁶⁰ The *Mirabile* court was primarily concerned with recognizing the role of the security interest exception in protecting secured creditors. This concern, however, caused the court to ignore the fact that the bank had acquired full legal title and not just a security interest. Instead, the court focused on whether the bank sufficiently participated in the management of the facility to remove it from the exception.¹⁶¹

153. *Id.*

154. *Id.*

155. *Id.*

156. *Id.*

157. *Id.*

158. *Id.*

159. *Id.*

160. United States v. *Mirabile*, 15 Envtl. L. Rep. (Env. L. Inst.) 20992, 20996 (E.D. Pa. Sept. 4, 1985).

161. *Id.*

The *Mirabile* court probably did not even consider the policy implications of its holding. In effect, the court's decision placed lenders in a more advantageous position at a sheriff's sale than other potential buyers.

The courts in *Maryland Bank* and *Guidice* approached the issue differently. These courts held that when a lender forecloses and acquires full title, it is removed from the security interest exception. The *Maryland Bank* court recognized the public policy ramifications of the *Mirabile* decision and stated that the *Mirabile* standard would allow foreclosing lenders to be unjustly enriched. For example, when a bank forecloses on a polluted property, it would be the only bidder at the sheriff's sale who could take title without being liable for the site's clean-up costs.¹⁶² Other potential buyers of the property would shy away from purchasing the contaminated property because they would incur liability for the clean-up costs of the property when they took title. Thus, the bank could purchase the property at a lower price, wait until the property has been cleaned up at the taxpayers' expense, and then sell the property at a significant profit once it becomes marketable.¹⁶³ The *Maryland Bank* court refused to approve of such a result. Instead, it held that when a lender forecloses and acquires title to property, it should be treated like any other bidder at the sheriff's sale and should incur liability the same as any other PRP.¹⁶⁴

The *Maryland Bank* position is and should be the dominant view espoused. As courts have recognized, both the public policy and structure of CERCLA support removing a mortgagee-turned-owner from the protection of the security interest exception. CERCLA was never intended to allow banks to profit from government funded clean-ups of contaminated property. To treat any subsequent purchaser of contaminated property differently under CERCLA is simply inequitable.

Legislative intent supporting this narrow interpretation of the security interest exception is evidenced by several factors. First, Congress allowed liability under CERCLA to be extremely widespread. This is evidenced by Congress imposing liability on four broadly defined classes of persons.¹⁶⁵ Second, Congress provided for only three meager defenses that are available to PRPs.¹⁶⁶ Finally, when Congress amended CERCLA in 1986, it did not significantly narrow any of CERCLA's liability provisions and did not alter any previous court interpretations of its provisions.¹⁶⁷

162. *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 573, 580 (D. Md. 1986).

163. *Id.*

164. *Id.*

165. See *supra* note 31 and accompanying text.

166. See *supra* note 83 and accompanying text.

167. See *supra* note 13 and accompanying text.

Even under the *Maryland Bank* view, lenders have an adequate means of protecting themselves from CERCLA liability. Financial institutions are in a position to investigate and discover potential environmental problems on their secured properties.¹⁶⁸ Additionally, banks can protect themselves by conducting reasonable investigations of potential borrowers' environmental situations *before* they lend money.

Recently, the *Guidice* court followed the *Maryland Bank* court's decision and entrenched its interpretation of the security interest exception into modern case law. It seems likely that the *Maryland Bank* position is now the dominant view because only one court has followed the *Mirabile* position after CERCLA was amended in 1986.¹⁶⁹

As a result of the *Maryland Bank* and *Guidice* decisions, lenders should be more cautious about foreclosing on their secured interests. In order to protect themselves, lenders must educate themselves about the hazardous waste disposal practices of their potential borrowers, conduct environmental inspections *before* making loans, and monitor the borrower's activities *during* the period of the loan. If after such investigation, a lender finds that a defaulting borrower has improperly disposed of hazardous substances on the property, the lender may be presented with a dilemma. The lender will have to decide: (1) if it is more cost effective to foreclose, purchase the property, pay for the clean-up, and resell the property or (2) if it is more cost effective to simply write-off the loan. Because of the staggering costs of environmental clean-up, lenders in many cases will find that it is better to simply write-off the loan.¹⁷⁰

B. Lenders Participating In A Borrower's Management

When Congress drafted the security interest exception, it did not define what actions constitute "participating in management." Consequently, this definition was left to judicial interpretation. Although substantial litigation has attempted to define the participation in management term, no clear standard or definition has emerged to provide

168. United States v. Maryland Bank & Trust Co., 632 F. Supp. 573, 580 (D. Md. 1986).

169. Coastal Casting Serv., Inc. v. Aron, No. 86-4463, 1988 WL 35012 (S.D. Tex. Apr. 8, 1988) (foreclosing on property is not enough in itself to find the lender liable as an owner).

170. In a recent survey of 2000 banks conducted by the American Banking Association, 15% of the banks reported that they had abandoned property held as collateral on loans, rather than foreclose upon it. The Financial Times Ltd., *Banks Criticize Proposed Lender Liability Rule*, WORLD INS. REP. (July 19, 1991). See also *Chemical Waste Complicates Many Land Sales, Financing*, Wall St. Journal, Nov. 5, 1986, at 39, col. 1 (lender decided not to foreclose on a \$200,000 loan upon learning that the land was contaminated with a hazardous substance and clean-up costs could amount to \$2.5 million).

guidance to lenders. Instead, there are two contrasting definitions of what constitutes permissible participation in a borrower's affairs before a lender is removed from the security interest exception. Under one view, courts have made a distinction between a lender's participation in the general financial management of a borrower's activities and a lender's impermissible participation in the day-to-day management of a borrower's production, operational, or waste disposal activities. This view was first outlined in *United States v. Mirabile* and has subsequently been followed by a line of district court opinions.¹⁷¹ This interpretation permits a lender to provide general *financial* advice and isolated instances of specific operational advice to a debtor without incurring CERCLA liability as long as a lender does not participate in the *day-to-day operational* aspects of the facility.¹⁷²

Although no appellate court has specifically espoused the *Mirabile* position, the Ninth Circuit in *In re Bergsoe Metal Corp.* seemed to implicitly adopt this view by holding that as a threshold matter, a lender must exercise *actual management authority* before it can be held liable for action or inaction that results in the discharge of hazardous wastes.¹⁷³ Furthermore, the court stated that a lender “[m]erely having the power to get involved in management, but failing to exercise it, is not enough” to subject it to liability.¹⁷⁴ This dicta should be interpreted as implicitly rejecting the *Fleet Factors* view which imposes liability on a lender whose actions merely indicate its *capacity* to become involved in a debtor's disposal decisions. Significantly, the *Bergsoe* court's treatment of the Port as a “secured lender” rather than an “owner” for purposes of CERCLA is also more consistent with the *Mirabile* position. Although the Port was the “technical owner” of the contaminated property, the *Bergsoe* court dismissed this fact and allowed the Port to claim protection under the security interest exception because it *acted* as a secured lender.¹⁷⁵

Contrary to the interpretation outlined in *Mirabile*, the Eleventh Circuit in *United States v. Fleet Factors Corp.* advocated a more encompassing definition of what activities may be classified as “participating in management” of a facility. In *Fleet Factors*, the court shifted the

171. See *United States v. Nicolet, Inc.*, 712 F. Supp. 1193, 1204-05 (E.D. Pa. 1989); *United States v. New Castle County*, 727 F. Supp. 854, 866 (D. Del. 1989); *Guidice v. BFG Electroplating & Mfg. Co.*, 732 F. Supp. 556, 561-62 (W.D. Pa. 1989); *Rockwell Int'l Corp. v. I.U. Int'l Corp.*, 702 F. Supp. 1384, 1390 (N.D. Ill. 1988); *United States v. Fleet Factors Corp.*, 724 F. Supp. 955, 960 (S.D. Ga. 1988), *aff'd on other grounds*, 901 F.2d 1550 (11th Cir. 1990).

172. See *supra* note 171.

173. *In re Bergsoe Metal Corp.*, 910 F.2d 668, 673 n.3 (9th Cir. 1990).

174. *Id.*

175. *Id.* at 671.

focus of its examination from the *type* of lender involvement (*i.e.*, financial or operational) to the *degree* of lender involvement, regardless of its type. Consequently, the court held that a lender will be removed from the security interest exception by participating in the financial or operational management of a facility to a degree indicating its capacity to influence the debtor's hazardous waste disposal decisions if it chose to do so.¹⁷⁶

The *Fleet Factors* decision established a third category of lender liability under CERCLA. Previously, lenders who foreclosed and acquired a facility or participated in the day-to-day operational management of a facility were subject to liability.¹⁷⁷ After *Fleet Factors*, secured lenders who merely have the capacity to influence a debtor's hazardous waste disposal decision are also potentially liable parties.

The *Fleet Factors* standard clearly narrows the scope of the security interest exception. It is relatively unclear, however, what lender management activities are permissible under *Fleet Factors*. The court offered some guidance by stating that lenders could still "monitor" any aspect of a debtor's business without incurring liability.¹⁷⁸ Furthermore, the court remarked that a lender "can become involved in occasional and discrete financial decisions relating to the protection of its security interest without incurring liability."¹⁷⁹ Lenders, however, should not heavily rely upon these gratuitous remarks. The consequences are severe for misjudging whether a certain action falls into one of these exceptions.

After examining the two differing interpretations of the security interest exception, it is suggested that the interpretation of "participating in management" advocated by *Mirabile* and several other courts better comports with the purpose and intent of CERCLA than the one proposed in *Fleet Factors*. First, the vagueness of the legal standard proffered by the *Fleet Factors* interpretation poses a significant problem for the lending community. The *Fleet Factors* standard imposes liability based on the *degree* of a lender's participation in a borrower's affairs. The court offered virtually no guidance as to what amount of participation is permissible. It only remarked that lenders could monitor any aspect of a borrower's financial situation and could become involved in "occasional and discrete financial decisions relating to the protection of its security interest."¹⁸⁰ This standard will result in lenders becoming overly cautious

176. *United States v. Fleet Factors Corp.*, 901 F.2d 1150, 1557-58 (11th Cir. 1990).

177. *United States v. Maryland Bank & Trust Co.*, 632 F. Supp. 578, 580 (D. Md. 1986). *See supra* note 171 and accompanying text.

178. *Fleet Factors*, 901 F.2d at 1558.

179. *Id.*

180. *Id.*

in making loans or engaging in loan workouts.¹⁸¹ In turn, this may cause a significant decrease in the availability of affordable credit to both small businesses and home purchasers whose property is environmentally active or near environmentally active property.¹⁸²

Furthermore, under the *Fleet Factors* standard, a lender may have no refuge from CERCLA liability if it conducts its business prudently and takes reasonable measures to protect its security interests. Virtually all lenders have the "capacity to influence," a borrower's business by the very terms of typical loan documents, security agreements, and mortgages. Without this "capacity to influence," a lender has no way to protect its collateral without incurring environmental liability. Therefore, by imposing liability on secured lenders who merely have the capacity to influence the environmental decisions of debtors, the *Fleet Factors* standard essentially eviscerates CERCLA's security interest exception.

Contrary to the *Fleet Factors* standard, the *Mirabile* standard imposes a palpable and discernable interpretation of the security interest exception. The *Mirabile* interpretation imposes liability based on a lender's *type* of participation (*i.e.*, financial or operational involvement), not its *degree* of participation. Although in some cases determining whether a lender is participating in the financial management or the operational management of a debtor may be difficult, the *Mirabile* standard provides lenders with predictability. Because of this increased predictability, this standard will not inhibit financial transactions unnecessarily, but will instead encourage responsible behavior and reasonable precautionary measures by lenders.

Second, the *Fleet Factors* interpretation is deficient because it ignores the fact that lenders often take actions to protect a security interest in a debtor's personal property as well as its real property. Under the *Fleet Factors* standard, a lender who does not have a secured interest in a borrower's real property may incur liability by merely becoming too involved in protecting an interest in a borrower's *personal* property. Lenders seldom conduct costly environmental investigations before mak-

181. In a recent survey of 2000 banks conducted by the American Banking Association, 63% of the banks reported that they had rejected loan applications because of possible environmental liability. Nearly 45% of the banks also reported that they had stopped lending to certain types of businesses, such as gasoline stations, because of potential environmental liability. The Financial Times Ltd., *Banks Criticize Proposed Lender Liability Rule*, WORLD INS. REP. (July 19, 1991).

182. Note, *Lender Liability for Hazardous Waste: An Economic and Legal Analysis*, 59 U. COLO. L. REV. 659, 669 (1988) [hereinafter Note, *Lender Liability*]. See also Leland, *Lender Liability and Cleanups Present Workout Dilemma*, AM. BANKER, June 20, 1990, at 4.

ing loans which are only secured by a borrower's accounts receivable, equipment, and inventory. The *Fleet Factors* interpretation, however, will force lenders to conduct these environmental audits on loans which are only secured by personal property.

Consequently, in order to protect themselves, lenders will need to conduct expensive environmental audits before making *any* type of secured loan. Small businesses seeking loans are hurt by such a result because the debtor must ultimately pay for the costs of these environmental audits. For example, a small business applying for a \$200,000 loan cannot afford to pay a \$10,000 to \$20,000 environmental audit fee.¹⁸³

Third, the *Fleet Factors* interpretation increases the likelihood that lenders will deny credit to businesses that either use hazardous substances or are located in areas of possible contamination. Already, small businesses such as dry cleaners, printers, pest control firms, agribusiness, and real estate developers are experiencing a credit crunch resulting from the lending community's attempts to avoid environmental liability.¹⁸⁴ Furthermore, besides tightening the availability of credit for developers and small businesses, the *Fleet Factors* standard will also make available credit more expensive.¹⁸⁵ This result will unduly burden many smaller businesses which depend upon the availability of affordable credit for their existence.¹⁸⁶ Finally, potential home purchasers will be denied credit merely because their property is located near an environmentally active facility.¹⁸⁷ CERCLA was not designed to chill responsible lending to potential home purchasers and environmentally active businesses.

Contrary to the *Fleet Factors* standard, the *Mirabile* standard effectuates a result in which lenders are not as prone to deny credit to businesses solely because of a business's affiliation with hazardous substances. Banks are better able to protect themselves from incurring liability under the lower and more predictable *Mirabile* standard. Consequently, the *Mirabile* interpretation is less burdensome on responsible parties seeking credit than the *Fleet Factors* interpretation.

183. Note, *Lender Liability*, *supra* note 182, at 669.

184. *FDIC, RTC Claim Federal Liability Laws Hamper Selling Insolvent Institutions*, 56 Banking Report (BNA) No. 15, at 685 (Apr. 15, 1991).

185. *Bankers, Lawyers, Business Endorse LaFalce CERCLA Lender Liability Bill*, 54 Banking Report (BNA) No. 23, at 986 (June 11, 1990) (testimony of Robert Clark, counsel for Associated Builders & Contractors, Inc. of Indiana, before the House of Representatives Small Business Committee). See also Note, *Lender Liability*, *supra* note 182, at 669.

186. Testimony of Lee Schroeder, President of First National Bank of Dana, Indiana, before the House of Representatives Small Business Committee (June 7, 1990) (unpublished) (on file in the office of the Indiana Law Review).

187. *Id.*

Finally, the *Fleet Factors* court asserted that its interpretation encourages lenders to monitor closely the hazardous waste policies and disposal practices of their debtors and requires their compliance with acceptable treatment standards as a prerequisite for continued financial support.¹⁸⁸ Facially, this seems to be a logical proposition. In reality, however, the *Fleet Factors* standard places lenders squarely on the horns of a dilemma. For example, if despite its previous investigations, a secured lender discovers a hazardous waste disposal problem on a defaulting debtor's site, the lender is faced with three options.

First, the lender may opt to foreclose. This is not usually a viable option, however, because the contaminated property will clearly not draw a good price at the foreclosure sale when any prospective purchaser (either a third party or the lender) will be subject to full liability upon acquiring title.

Second, upon the debtor's default, the lender may choose to call the entire loan due, obtain what money it can from the debtor, and then write-off the remaining loss. In doing so, the lender will only minimally inform the debtor of its financial situation because the lender will not want its actions to be misconstrued so that it becomes ensnared in the *Fleet Factors* web of liability. This result defeats the purposes of CERCLA. CERCLA's liability scheme was designed not only to impose liability upon responsible parties, but also to *encourage* activities that lead to safe hazardous waste disposal practices. In this situation, the *Fleet Factors* standard actually *discourages* responsible activities because it gives lenders incentives to terminate rather than to continue their association with a financially troubled debtor.¹⁸⁹ This termination will often cause the debtor to become insolvent. Thus, the government will in many cases be left with the responsibility to clean-up the contaminated site because the property will be virtually unmarketable.

Third, as the court in *Fleet Factors* suggested, a lender may attempt to engage in an active workout agreement with a defaulting debtor that attempts to address the hazardous waste disposal problems at the facility. In doing so, however, a lender probably subjects itself to liability under the *Fleet Factors* standard because its actions may be deemed of a sufficient degree to support the inference that it could have an effect on hazardous waste disposal decisions if it so chose. Because lenders may be held liable regardless of the success of their efforts, lenders will be more apt *not* to engage in active workout arrangements with debtors. Again, this is an undesirable result because it discourages good faith

188. United States v. Fleet Factors Corp., 901 F.2d 1550, 1558-59 (11th Cir. 1990).

189. Guidice v. BFG Electroplating & Mfg. Co., 732 F. Supp. 556, 562 (W.D. Pa. 1989).

workout arrangements designed to address environmental problems.

Under the *Mirabile* standard, lenders are not discouraged, but rather are encouraged to engage in good faith workout arrangements designed to address a borrower's environmental problems. The *Mirabile* interpretation allows secured lenders to participate in a debtor's management decisions without incurring liability as long as the secured lender does not participate in the day-to-day operational decisions of the company. By encouraging structured workout arrangements, this standard increases the likelihood that borrowers will remain solvent and will consequently be able to undertake, fund, and remedy part, if not all, of their environmental problems.

Furthermore, the *Mirabile* standard encourages potential borrowers to engage in environmentally safe practices before they apply for loans. Borrowers will engage in such practices because they know that lenders will conduct environmental audits and will structure the loan agreement terms according to the environmental liability risks present. Thus, both lenders and borrowers have incentives to require and promote environmentally safe practices. The *Guidice* court summarized the policy ramifications of the two competing standards by stating:

A goal of CERCLA is safe handling and disposal of hazardous waste. To encourage banks to monitor a debtor's use of security property, a high liability threshold [such as the *Mirabile* position] will enhance the dual purposes of protection of the banks' investments in promoting CERCLA's policy goals. Conversely, a low liability standard [such as the *Fleet Factors* position] would encourage a lender to terminate its association with a financially troubled debtor and expedite loan payments in an effort to recover the debts.¹⁹⁰

The reasoning in *Guidice* is sound, and its holding should be espoused by other courts. The *Mirabile* standard best comports with the purpose, intent, and underlying policy interests of CERCLA.

To the extent that the *Fleet Factors* decision is cited to establish a new standard of lender liability, it should be viewed as an aberration that will be rejected by other courts that are more concerned with the public policy ramifications of this issue. To the extent that *Fleet Factors* is limited to its facts, however, the decision is entirely consistent with the trend of authority. In *Fleet Factors*, there was sufficient active involvement in the day-to-day operational management of the debtor's facility so that the lender would probably be held liable under either the *Fleet Factors* or *Mirabile* standards of liability. Consequently, courts

190. *Id.* (citations omitted).

can and should distinguish the *Fleet Factors* decision on the basis of its facts and refuse to follow its holding. Courts should instead espouse the *Mirabile* standard because it comports with both the policies and intent of CERCLA.

C. Proposed EPA Rule Interpreting The Security Interest Exception

In response to the clamor of uncertainty caused by the expansive liability standard set forth in the *Fleet Factors* decision, the EPA recently promulgated proposed regulations which seek to provide secured lenders with a safe harbor from CERCLA liability.¹⁹¹ These regulations are designed to specify actions that lenders may take while making loans, during the life of loans, during workouts, and upon foreclosure, while still remaining within the bounds of the security interest exception. If these regulations become effective, however, they will provide little relief to foreclosing lenders and lenders who would otherwise incur liability under the *Fleet Factors* standard.

After its first draft rule was widely criticized, the EPA submitted a second draft rule to the Office of Management and Budget for review.¹⁹² This rule can essentially be divided into two main components. First, the proposed rule addresses lender activity that constitutes “participation in management” for purposes of applying the security interest exception.¹⁹³ The draft rule states that a lender has “participated in management” only if, while the borrower was still in possession, the secured lender “materially divested the borrower of decision making control over vessel or facility operations, particularly with respect to the hazardous substance present at the vessel or facility.”¹⁹⁴ Significantly, the proposed rule also attempts to regulate the *Fleet Factors* standard out of existence by stating that “participation in management . . . does not include the mere unexercised capacity or ability to influence vessel or facility operations.”¹⁹⁵

In addressing the permissible scope of a lender’s activities, the proposed rule allows lenders to protect secured interests by policing loans or by undertaking financial workouts with borrowers when its security interests are threatened. The rule also provides a list of examples and

191. National Oil and Hazardous Substances Pollution Contingency Plan; Lender Liability under CERCLA, 56 Fed. Reg. 28798-28810 (1991) (to be codified at 40 C.F.R. § 300.1100).

192. *Id.* At the time of publication, the EPA proposed rule was undergoing public comment and judicial review. The proposed rule becomes effective in October 1991 if it satisfactorily completes this process.

193. *Id.* at 28809 (to be codified at 40 C.F.R. § 300.1100(c)(1)).

194. *Id.*

195. *Id.*

illustrations of activities that a lender may undertake without losing the security interest exemption provided that the actions are necessary to protect the security interest.¹⁹⁶ Specifically, the rule provides that a secured creditor may:

1. Require clean-up of a facility before granting a loan or during the life of a loan;
2. Require the facility owner or operator to give assurances of compliance with applicable environmental laws;
3. Periodically inspect or regularly monitor the facility and the owner or operator's business or financial condition;
4. Require or conduct environmental audits prior to or during the period of the loan;
5. Provide periodic financial advice to a debtor;
6. Engage in loan workout activities, including restructuring or renegotiating the loan, requiring payment of additional interest, extending the payment period, exercising forbearance or providing advice, or take other action that is necessary to protect the security interest;
7. Wind up operations, liquidate assets or otherwise act to recover the value of the collateral in a manner consistent with good commercial and environmental practice; and
8. Impose other requirements reasonably necessary for the lender to police the loan adequately or to comply with legal requirements.¹⁹⁷

The EPA expressly considers these activities to be consistent with the security interest exception and does not consider them to constitute "participation in management" for purposes of the security interest exception. The EPA's proposed rule requires lenders to duly consider and account for the hazardous substances known to be present at a facility. Lenders who act or fail to act in a way that causes or contributes to contamination will be excluded from the security interest exception.¹⁹⁸

The second component of the proposed EPA rule deals with secured lenders that foreclose and acquire title to an interest.¹⁹⁹ The draft regulation will allow secured lenders to foreclose and acquire title without incurring environmental liability as long as the lender disposes of the property within a reasonable period of time.²⁰⁰ If the lender divests the foreclosed property within six months of taking title, the lender is

196. *Id.* (to be codified at 40 C.F.R. § 300.1100(c)(2)).

197. *Id.* (to be codified at 40 C.F.R. §§ 300.1100(b)(1)-(c)(2)).

198. *Id.* (to be codified at 40 C.F.R. § 300.1100(c)(1)).

199. *Id.* at 28808-09 (to be codified at 40 C.F.R. § 300.1100(b)(1)).

200. *Id.* at 28809 (to be codified at 40 C.F.R. § 300.1100(b)(1)(ii)).

presumed to fall under the protection of the security interest exception.²⁰¹ If the lender does not divest itself of the property within the six month time frame, then the burden shifts to the lender to demonstrate that its actions were designed primarily to protect the security interest and were not designed for investment purposes.²⁰² This component of the rule seeks to nullify the holdings in *Guidice* and *Maryland Bank* that lenders who foreclose and acquire title become PRPs and incur CERCLA liability.

Although the EPA proposed rule places welcome limits on the government's efforts to impose liability on secured lenders, it does not apply to third party contribution actions against secured creditors.²⁰³ When the EPA is a plaintiff in an environmental lawsuit, the proposed EPA rule will be predictive of the agency's actions and how the EPA will view the actions of secured lenders. In private third party contribution lawsuits in which the secured creditor is the defendant, however, the proposed EPA rule will not provide protection to the secured lender because it is merely an interpretive rule which is not judicially binding.²⁰⁴

Because the proposed rule is not judicially binding, it offers virtually no protection to secured creditors. In nearly all CERCLA cases, there are multiple PRPs for every site. Thus, even if the EPA does not file suit against a secured creditor because it is following the EPA rule, one of the other PRP defendants (usually the present owner or operator of the facility) will bring a third party contribution action against the secured lender.²⁰⁵ The EPA rule will not apply to this third party action, however, and a court will be free to apply the *Mirabile* standard, the *Fleet Factors* standard, or some other liability standard in its examination of the lender's activities. If this occurs, lenders may incur liability despite the existence of the EPA rule. Consequently, although the EPA's proposed rule may be useful for secured lenders in direct actions brought by the EPA, it provides little or no assistance to secured lenders in third party contribution actions when the secured creditor is the defendant.

201. *Id.*

202. The proposed rule does not apply the six month limitation to government lenders and institutions that foreclose on property. These entities are presumed to be holding foreclosed property primarily to protect a security interest and for investment purposes. *Id.*

203. National Oil and Hazardous Substances Pollution Contingency Plan; Lender Liability Under CERCLA, 56 Fed. Reg. 28798-28810 (1990) (to be codified at 40 C.F.R. § 300.1100).

204. Because the EPA rule is interpretive in nature and is a significant departure from past EPA policy, courts will generally hold that such an interpretation should be given less deference than a relatively consistent line of court decisions.

205. For example, the following cases each consisted of third party suits against secured lenders: *In re Bergsøe Metal Corp.*, 910 F.2d 668 (9th Cir. 1990); *Guidice v. BFG Electroplating & Mfg. Co.*, 732 F. Supp. 556 (W.D. Pa. 1989); *United States v. Mirabile*, 15 Envtl. L. Rep. (Envtl. L. Inst.) 20992 (E.D. Pa. Sept. 4, 1985).

IV. PRECAUTIONARY MEASURES FOR LENDERS

Lenders should not wait until a borrower defaults on a loan before taking measures to prevent the imposition of liability costs under CERCLA. Instead, lenders must anticipate environmental problems that may arise under CERCLA and institute policies and procedures which address the potential risks of environmental liability.

First, the lender can protect itself from liability by insisting that the mortgagor obtain Environmental Impairment Liability (EIL) Insurance.²⁰⁶ EIL policies are designed to cover claims arising from environmental damage to the covered property. Although these policies may afford lenders some protection, lenders should not rely solely on EIL policies for protection. Currently, no EIL policy completely covers a lender's potential CERCLA exposure.²⁰⁷ In addition, mortgagors who have defaulted on their loans are also likely to forego payments on their EIL policies. This will cause the policies to lapse, and lenders will lose their protection.

Second, a lender should conduct an environmental audit and risk assessment on the potential debtor's property *before* making loans. Environmental audits are the best way that lenders can minimize CERCLA liability risks. It is recommended that an environmental investigation consist of a two-phase approach which includes: (1) a review of readily ascertainable environmental information concerning the property and (2) a scientifically conducted environmental audit to detect hazardous waste contamination on the property.²⁰⁸

In phase one, the lender should review state and federal government records of the particular property to ascertain whether the present or prior owners have violated any environmental laws, permits, or other requirements. In addition, lenders may use CERCLIS, the Superfund Data Base, to find out whether a secured property is located within a one mile radius of a potentially dangerous hazardous waste site.²⁰⁹ Because phase one activities do not require special skills, the investigation can be conducted by bank personnel inexpensively.²¹⁰

In phase two, a specialist should be hired to investigate both the surface and subsurface of the property. This audit should cover the

206. See Gieser, *Federal and State Environmental Law: A Trap for the Unwary Lender*, 1988 B.Y.U. L. REV. 643, 694.

207. *Id.*

208. The two-step technique was developed by the Federal National Mortgage Association. See Comment, *Buying into Trouble—Lender Liability under CERCLA and SARA*, 14 S. ILL. L.J. 319, 336 (1990).

209. *Id.* at 337.

210. The cost for a typical phase one audit can range from \$1500 to \$7500. *Id.* at 337 n.128.

property's soil and water samples, machinery, inventory, electrical systems, and construction materials. Phase two audits can be costly,²¹¹ but the cost of these audits is small compared to a potential multimillion dollar clean-up liability.

Third, the lender should require the borrower to declare that no past or current user of the property has created conditions which may give rise to environmental liabilities and that no enforcement actions are pending. In addition, to the extent that any environmental problems are discovered during the audit, those problems should be addressed *prior* to closing. The mortgagor should be required to clean-up any environmental hazards and comply with all present and future environmental laws.²¹²

Fourth, lenders should include an indemnity clause in loan agreements in which the mortgagor is required to indemnify the lender for clean-up costs upon the discovery of environmental contamination. Indemnity agreements have substantial problems, however. An indemnity agreement is only useful if the borrower is solvent, and in most cases, mortgagors in default will not be solvent. Additionally, indemnity agreements do not protect lenders from liability under CERCLA, but only provide a means for lenders to recoup their response costs from the mortgagor.²¹³

Fifth, lenders can minimize their potential liability by using contractual clauses that require a borrower's environmental compliance. Loan agreements should allow the lender to closely monitor the borrower's compliance with its environmental covenants. Additionally, a provision should be added requiring the borrower to notify the lender of any environmental problems. It is suggested that these provisions be tailored to closely follow the provisions of the EPA draft rule, especially with respect to the lender's ability to participate in management. It is also important that the lender include an acceleration clause that allows the lender to call the entire note due when the borrower fails to comply with the other covenants. This provision may allow the lender to recoup its loan proceeds before any serious environmental liability risks develop.²¹⁴

Sixth, lenders must be extremely careful not to become overly involved in a debtor's financial or operational management. Although isolated and discrete involvement may be safe, lengthy negotiations and enforcement steps designed to bring a borrower into compliance with loan provisions may remove the lender from the security interest exception

211. See Gieser, *supra* note 206, at 699 n.248. Environmental testing can cost up to \$20,000-\$30,000. *Id.*

212. *Id.* at 700.

213. 42 U.S.C. § 9607(e) (1988).

214. See Klotz & Kane, *supra* note 33, at 37-38.

and expose it to significant liability. Consequently, lenders should not engage in active workout arrangements with defaulting debtors unless the potential clean-up costs are low.

Seventh, lenders should weigh foreclosing decisions carefully. Put simply, if a high risk of incurring significant liability costs upon the acquisition of title exists, a lender should not foreclose. Instead, the lender should write-off the loan or engage in a workout arrangement in which the lender has only limited involvement with the borrower's management.

Finally, lenders should closely monitor the developments in both state and federal laws. Currently, several states have adopted "superlien" statutes which allow the imposition of priority liens on polluted properties.²¹⁵ In addition, Congress is presently considering bills that will provide different types of exemptions from CERCLA liability to lending institutions.²¹⁶ If passed, these bills will effectively nullify the results reached in *Fleet Factors*, *Guidice*, and *Maryland Bank*.

V. CONCLUSION

Judicial interpretation of CERCLA's security interest exception has had a checkered history and result. Although courts seem to agree that lenders who acquire title through foreclosure of a secured interest are subject to liability, there are conflicting standards as to what activities remove lenders from the protection of the security interest exception.

Under the line of district court cases espousing the *Mirabile* standard, a lender is subject to liability only through its participation in a borrower's day-to-day operational activities. This standard encourages a lender to monitor a borrower's activities and to become involved in the debtor's financial management in order to protect its security interest. By encouraging workout arrangements, this standard promotes CERCLA's goals by reducing the probability that borrowers will become insolvent and leave the government with enormous response costs.

215. Note, *Connecticut Lender Liability under Federal and State Environmental Law*, 9 BRIDGEPORT L. REV. 175, 176-77 (1988).

216. H.R. 1450, 102d Cong., 1st Sess., 137 CONG. REC. 987 (1991). This bill provides that a secured lender's participation in a company must be concrete in order for the lender to be liable for contamination. Lenders are also excluded from liability during a workout or foreclosure. Lenders would still incur liability if they caused or added to the environmental problem, however. *Id.* See also S. 651, 102d Cong., 1st Sess., 137 CONG. REC. 3457 (1991). This bill would place a cap on liability for insured depository institutions for environmental contamination in cases in which the lending institution was not at fault. The liability limit would not apply to lending institutions that cause or add to the environmental problem, however. S. 651, 102d Cong., 1st Sess., 137 CONG. REC. 3457 (1991).

The *Fleet Factors* standard on the other hand, is an aberration in the judicial interpretation of the security interest exception. Under this standard, a lender may incur liability merely by becoming overly active in the borrower's financial or operational management. This standard is not only vague and unpredictable, but is also contrary to CERCLA's purpose. In reality, this standard encourages lenders to terminate their relationship with defaulting debtors who possess contaminated properties, rather than engage in active workout arrangements designed to alleviate the problems. Consequently, courts should limit the *Fleet Factors* decision to its facts and espouse the *Mirabile* standard.

Clearly, lenders can take affirmative steps to protect themselves from CERCLA liability. Lenders should conduct environmental audits and include provisions in loan agreements allowing them to monitor a borrower's compliance with accepted environmental standards. If, despite all due diligence, a secured lender discovers that the borrower's property is contaminated, only with great caution should the lender exert any significant control over the debtor's financial or operational decisions regarding the facility.

Regardless of whether courts adopt the *Mirabile* or *Fleet Factors* interpretation of the security interest exception, one thing is clear. Lenders now are faced with three concerns: credit, collateral, and clean-up.

SCOTT R. ALEXANDER

Statutory Control of DNA Fingerprinting in Indiana

INTRODUCTION

In 1986, a woman was raped in a small town in Leicestershire, England. The only evidence of the perpetrator was a semen sample recovered from the victim. The police were aware that a local scientist, Alec Jeffreys,¹ was developing a technique for comparing deoxyribonucleic acid (DNA) samples that might be used for identification purposes. The police asked Dr. Jeffreys to test DNA samples from each of the four thousand men in the town. One of the persons to be tested, Colin Pitchfork, engaged a substitute to take the test for him. The police discovered this attempted subterfuge and arrested Pitchfork. Pitchfork later confessed to the rape, thereby earning the dubious distinction of being the first person to be convicted by DNA fingerprinting evidence.²

DNA fingerprinting has been described as the most important advance in criminology since the advent of cross-examination.³ In common with traditional genetic analyses such as ABO typing, human leucocyte antigen typing, and typing of red cell enzymes and serum proteins,⁴ DNA fingerprinting serves to indicate whether two samples of human tissue, one recovered from the scene of the crime and the other from a suspect, share common characteristics. DNA fingerprinting, however, differs from traditional techniques in that it requires a smaller sample size and it gives rise to a much higher exclusion frequency.⁵

DNA fingerprinting was first used in the United States in 1987, one year after Pitchfork's conviction.⁶ As of October 1990, it has been used in 2000 criminal investigations and 200 trials in thirty-eight states.⁷ It has also been used in thousands of civil cases, principally paternity

1. Alec Jeffreys is a professor of biochemistry at Leicester University, England.

2. See Moss, *DNA—The New Fingerprints*, 74 A.B.A. J. 66 (1988).

3. People v. Wesley, 140 Misc. 2d 306, 308, 533 N.Y.S.2d 643, 644 (1988).

4. For a discussion of these analyses, see P. GIANNELLI & E. IMWINKELRIED, *SCIENTIFIC EVIDENCE* 565-632 (1986).

5. Exclusion frequency, as used herein, is the frequency with which characteristics shared by the two samples under test occur in the population at random. Exclusion frequencies as high as one in 30 billion have been claimed for DNA fingerprinting. Dodd, *DNA Fingerprinting in Matters of Family and Crime*, 318 NATURE 506, 506 (1985). Exclusion frequencies of about one in 10 to one in 100 are more typical for traditional genetic analyses. *Id.* at 507.

6. Slackman, *Genetic Finger-Pointing; Prosecutors Fear Impact of Surprise Bill Regulating DNA Evidence*, NEWSDAY, July 7, 1990, at 3.

7. Marcotte, *Report: DNA Tests Valid*, 76 A.B.A. J. 26, 26 (1990).

disputes.⁸ Most of these DNA tests were performed by two commercial laboratories, Cellmark and Lifecodes. Lifecodes now earns \$40 million each year from the process.⁹

In the initial case, Pitchfork's confession spared the court from having to assess the reliability of DNA fingerprinting evidence. Reliability has, however, been closely scrutinized in later cases. Such scrutiny is inevitable in light of the complexity of the DNA fingerprinting process and the possibility of conviction solely on the basis of DNA fingerprinting evidence.¹⁰ In 1989, two courts refused to admit DNA fingerprinting into evidence.¹¹ These courts accepted that the principles underlying DNA fingerprinting are sound, but found that the particular procedures of the laboratories conducting the test were unreliable.¹² More recent appellate decisions have generally affirmed trial court decisions to admit DNA fingerprinting into evidence and have sustained both the fundamental principles and the particular procedures of the laboratories conducting the tests.¹³

8. *Id.* This Note principally addresses the use of DNA fingerprinting in criminal, rather than civil cases. Reliability concerns are not as great in civil cases because the comparison is made between DNA samples derived from two people (e.g., alleged father and son in a paternity proceeding), rather than between one sample derived from the scene of a crime and another from a person, as in criminal applications. Unlike samples derived from the scene of the crime, samples derived from people are clean and can be as large as required. DNA fingerprinting can be performed with greater accuracy on large clean samples. A large sample also allows repetition of the test if the results are ambiguous. Paradoxically, the Indiana statute, which facilitates admission of DNA fingerprinting evidence, applies only to criminal trials in which reliability concerns are greatest.

9. Unger, *Court Challenge Casts Pall Over DNA Testing Industry*, NEWSDAY, July 30, 1989, at 47.

10. In *Spencer v. Commonwealth*, 240 Va. 78, 393 S.E.2d 609 (1990), the Virginia Supreme Court affirmed a capital conviction in which there was little evidence other than DNA fingerprints. A juror interviewed after an unpublished New York criminal trial commented, "DNA was kind of a sealer on the thing. You can't really argue with science . . . that was the whole case in my opinion." Note, *The Dark Side of DNA Profiling: Unreliable Scientific Evidence Meets the Criminal Defendant*, 42 STAN. L. REV. 465, 515 (1990) [hereinafter Note, *The Dark Side*].

11. *State v. Schwartz*, 447 N.W.2d 422 (Minn. 1989); *People v. Castro*, 144 Misc. 2d 956, 545 N.Y.S.2d 985 (1989).

12. *Schwartz*, 447 N.W.2d at 427; *Castro*, 144 Misc. 2d at 978, 545 N.Y.S.2d at 999.

13. *Martinez v. State*, 549 So. 2d 694 (Fla. Dist. Ct. App. 1989); *Andrews v. State*, 533 So. 2d 841 (Fla. Dist. Ct. App. 1989); *Caldwell v. State*, 260 Ga. 278, 393 S.E.2d 436 (1990) (approving the trial court's decision to admit DNA fingerprinting evidence, but remanding for retrial on the grounds that the testing laboratory overestimated the exclusion frequency); *Cobey v. State*, 73 Md. App. 233, 533 A.2d 944 (1989); *State v. Pennington*, 327 N.C. 89, 393 S.E.2d 847 (1990); *State v. Ford*, 301 S.C. 485, 392 S.E.2d 781 (1990); *Glover v. State*, 787 S.W.2d 544 (Tex. Ct. App. 1990); *Kelly v. State*, 792 S.W.2d 579 (Tex. Ct. App. 1990).

Indiana Code section 35-37-4-10 follows the trend of recent appellate court decisions toward increasing acceptance of DNA fingerprinting evidence. The statute provides: "In a criminal trial or hearing, the results of forensic DNA analysis are admissible in evidence without antecedent expert testimony that forensic DNA analysis provides a trustworthy and reliable method of identifying characteristics in an individual's genetic material."¹⁴ It is obvious that the statute is intended to make DNA fingerprinting evidence more easily admissible into Indiana courts by reducing the expert testimony that is required to establish a foundation of reliability. It is less clear, however, whether the statute is intended merely to recognize that reliable techniques for performing DNA fingerprinting can be devised, whether it is intended to endorse the particular procedures of currently recognized testing laboratories, or whether it is intended as a *per se* rule of admissibility.

This Note shows how these three interpretations arise when the statute is viewed in light of Indiana's common-law approach to admissibility of scientific evidence. This Note attempts to discern which interpretation was intended by the Indiana legislature by referring to case law in other jurisdictions, by analogizing to other statutes, and by tracing the policy consequences of each interpretation. This Note also considers possible constitutional challenges to the statute.

I. PRINCIPLES OF DNA FINGERPRINTING

DNA fingerprinting (otherwise known as DNA profiling, DNA typing, or forensic DNA analysis) is an analytical technique for detecting differences between DNA molecules. This technique is based on the fundamental principle that, between different people, there are differences between their respective DNA molecules that give individuals their unique character.¹⁵ A DNA molecule comprises millions of copies of four different units known as bases.¹⁶ The sequence of bases along the molecule forms a code that is different for each individual.¹⁷ Perhaps surprisingly, differences between the DNA sequences of different individuals amount to less than one percent of their total DNA codes.¹⁸ However, these differences are not distributed at random throughout the code, but are concentrated at several specific locations known as "polymorphic regions."¹⁹ At polymorphic regions, there is a high probability of finding

14. IND. CODE ANN. § 35-37-4-10(b) (Burns Supp. 1990) (effective July, 1990).

15. See generally J. WATSON, THE MOLECULAR BIOLOGY OF THE GENE (1987).

16. See L. STRYER, BIOCHEMISTRY 71-76 (3rd ed. 1988).

17. *Id.*

18. See Note, *The Dark Side*, *supra* note 10, at 470-71.

19. Polymorphic sites are also referred to as "minisatellites." See Jeffreys, *Highly Variable Minisatellites and DNA Fingerprints*, 15 BIOCHEMICAL SOC'Y TRANSACTIONS 309 (1987).

different sequences in different individuals. In DNA fingerprinting, a DNA sample from the scene of the crime is compared with a sample from a suspect at several polymorphic regions. If the two DNA samples are different at any of the polymorphic regions tested, the two samples are from different people. If the two DNA samples are identical at all polymorphic regions tested, the two samples are likely derived from the same person.²⁰

II. PROCEDURES FOR PERFORMING DNA FINGERPRINTING AND SOURCES OF ERROR

The original method for performing DNA fingerprinting, and one that has been the subject of the most litigation to date, is known generically as restriction fragment length polymorphism (RFLP) analysis. RFLP analysis is practiced by three laboratories: Lifecodes, Cellmark (both commercial), and the FBI. A third commercial laboratory, Cetus, uses a different method based on the polymerase chain reaction.²¹ A further method based on DNA sequencing is under development.²²

A. RFLP Analysis

1. Procedure.—RFLP analysis is a complex procedure that can be subdivided into eight individual steps.²³

a. DNA extraction

DNA is extracted from two samples. One sample is from human cells (e.g., semen, blood, or hair) found at the scene of the crime. The

20. *But see infra* note 29 and accompanying text.

21. The polymerase chain reaction is a recently discovered procedure for artificial reproduction of DNA. See Mullis, Erlich, Arnheim, Horn, Saiki, & Scharf, *Process for Amplifying, Detecting and/or Cloning Nucleic Acid Sequences*, U.S. Patent 4,683,195 (1987).

22. DNA sequencing is a technique for determining the sequence of bases comprising the DNA code. See L. STRYER, *supra* note 16, at 120-23.

23. The FBI, Lifecodes, and Cellmark procedures for performing RFLP analysis are respectively described in: Jakobetz v. United States, 747 F. Supp. 250, 251-54 (D. Vt. 1990); Baird, Balazs, Guisti, Miyazaki, Nicholas, Wexler, Kanter, Glassberg, Allen, Rubinstein, & Sussman, *Allele Frequency Distribution of Two Highly Polymorphic DNA Sequences in Three Ethnic Groups and Its Application to the Determination of Paternity*, 39 AM. J. HUM. GENETICS 489 (1986) [hereinafter Baird]; Jeffreys, Willson, & Thein, *Individual-Specific "Fingerprints" of Human DNA*, 316 NATURE 76 (1985) [hereinafter Jeffreys]. The description of the RFLP procedure given in this Note is schematic only and is based in part on these sources and in part on T. MANIATIS, E. FRITSCH, & J. SAMBROOK, *MOLECULAR CLONING: A LABORATORY MANUAL* (1982) [hereinafter T. MANIATIS]. A more detailed account is given by Thompson & Ford, *DNA Typing: Acceptance and Weight of the New Genetic Identification Tests*, 75 VA. L. REV. 45 (1989).

other sample is taken directly from the suspect (e.g., a blood sample). Subsequent steps are performed on the two samples in parallel.

b. DNA cleavage

The DNA is cleaved by use of enzymes known as restriction endonucleases that recognize and cut specific sequences of the DNA. This cleavage converts single molecules of DNA, each comprising millions of bases, to thousands of smaller fragments.²⁴ The vast majority of fragments generated from an individual's DNA will be identical to fragments derived from corresponding regions of another individual's DNA. However, a small proportion of fragments derived from polymorphic regions will differ in size from fragments derived from corresponding polymorphic regions in a different individual.²⁵ Subsequent steps in the procedure are designed to detect size differences between fragments derived from corresponding polymorphic regions in the two samples.

c. Electrophoresis

The DNA fragments are separated according to their size by a technique known as agarose gel electrophoresis. The procedure entails inserting DNA fragments into a gel and applying an electric field across the ends of the gel. DNA is negatively charged and hence, moves to the positive terminal. Small fragments are less impeded by the gel than large fragments and hence, move faster. Thus, after electrophoresis, the fragments are arranged in size from largest to smallest while immersed in the gel. Cleavage and separation of fragments allow analysis of polymorphic fragments free from the vast majority of fragments which exactly correspond in different individuals.

d. Southern transfer

The DNA fragments are transferred from the fragile gel to a more robust nylon membrane in a procedure known as Southern transfer.²⁶ The nylon membrane is layered directly over the gel and a stack of tissues is placed on top of the membrane. Fluid from the gel is drawn through the membrane into the tissues. DNA moves with the fluid from the gel, but comes to rest on the membrane to which it becomes permanently bound. Southern transfer preserves the separated array of fragments in the same relative positions as on the gel.²⁷

24. L. STRYER, *supra* note 16, at 130.

25. See Thompson & Ford, *supra* note 23, at 67-68.

26. Southern, *Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis*, 98 J. MOLECULAR BIOLOGY 503 (1973).

27. *Id.*

e. Hybridization

The nylon membrane containing the array of DNA fragments is immersed in a hybridization fluid containing a radioactive probe specific for a polymorphic region. The probe binds to the fragments containing this polymorphic region which are themselves immobilized to the nylon membrane.

f. X-ray photography²⁸

The nylon membrane is overlaid with film. The radioactive probe bound to polymorphic fragments exposes the film specifically at the positions where the polymorphic fragments have located. The polymorphic fragments derived from the two samples are compared to determine whether there is a match as judged by coincidence of band positions.

g. Rehybridization

The probe is washed off the nylon membrane and hybridization and X-ray photography (steps e and f) are successively repeated for two or three other probes. As an alternative to rehybridization, when sufficient DNA is available, the pair of samples under test may be split into fractions after DNA extraction (step a). Steps b - f are then performed separately for paired fractions and each pair is hybridized with a different probe at step e.

h. Statistical analysis

If a difference in polymorphic fragments is identified for any of the probes used, the two DNA samples are not identical and therefore, derive from different people. If the polymorphic fragments match for all of the probes used, an exclusion frequency is calculated. The exclusion frequency, as used herein, is the frequency with which the characteristics shared by the two DNA samples under test occur in the general population. If there is no independent incriminating evidence, the exclusion frequency is the relative probability that the defendant committed the crime compared with a person selected at random from the general population. The exclusion frequency is not, however, the same as the probability that the defendant committed the crime.²⁹

28. Technically known as "autoradiography."

29. If independent incriminating evidence exists that can be translated into a probability that the defendant committed the crime, this probability can be combined with the exclusion frequency to generate an overall probability of guilt using Bayesian analysis. See Finkelstein & Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970).

2. *Possible Sources of Error.*—The procedure outlined in the preceding section comprises many successive steps each of which is subject to a multitude of errors which can be subdivided into three categories.

a. *Anomalous fragments*

A match is declared by comparing the size of polymorphic fragments between two samples. Theoretically, the polymorphic fragments from each sample should be the only fragments visible on the photograph from which the comparison is made. Yet, a variety of errors can result in the appearance of extra spurious fragments or the disappearance of genuine polymorphic fragments. Either occurrence may make the comparison erroneous.

Spurious extra fragments can result from the addition of either too much enzyme (resulting in star activity and extra small fragments) or too little enzyme (resulting in partial digestion and extra large fragments).³⁰ Determining how much enzyme to add is difficult because the activity of the enzyme may decay with storage³¹ and because it is not known precisely how much DNA there is in the sample to be digested.³² Spurious fragments may also result from the contamination of a sample with bacterial DNA (this is particularly likely for DNA recovered from the scene of the crime)³³ or by cross-contamination between the two samples (which could arise from operator error).³⁴ Spurious bands can also result from nonspecific binding of the radioactive probe to fragments other than those from the polymorphic region under test.³⁵ The extent to which nonspecific binding causes problems depends on the hybridization conditions and the length of time the film is exposed.

30. Restriction endonucleases recognize and cut at specific sequences of the DNA code. If the right amount of enzyme is added, the enzyme cuts at each occurrence of its recognition sequence and at no other sequences. If too much enzyme is added, the enzyme may cut at sequences resembling its recognition sequence (referred to as "star activity"). If too little enzyme is added, the enzyme will not cut all of its recognition sequences present on the DNA. See Fuchs & Blakesley, *Guide to the Use of Type II Restriction Endonucleases*, 100 METHODS IN ENZYMOLOGY 3, 33-38 (1983).

31. The FBI claims to check the activity of each batch of enzyme before performing a test. See United States v. Jakobetz, 747 F. Supp. 250, 257 (D. Vt. 1990).

32. DNA concentration is easily measured provided that a sufficiently large sample of DNA is available. See, e.g., J. ZYSKIND & L. BERNSTEIN, RECOMBINANT DNA LABORATORY MANUAL 17-19 (1989). A possible difficulty with measuring DNA concentration in the course of performing DNA fingerprinting is that the available sample of DNA may be so small that to use up some of it in determining its concentration risks not having a sufficient remaining sample to perform the fingerprinting analysis.

33. People v. Castro, 144 Misc. 2d 956, 969, 545 N.Y.S.2d 985, 993 (1989); Lander, *DNA Fingerprinting on Trial*, 339 NATURE 501, 503 (1989).

34. Thompson & Ford, *supra* note 23, at 95.

35. See Jeffreys, *supra* note 23, at 77 (legend to Fig. 1).

Not only can spurious extra fragments appear, but genuine polymorphic fragments can also disappear for a variety of reasons. One such reason is the possibility of exonuclease contamination.³⁶ All procedures involving DNA are performed by technicians wearing plastic gloves who use sterile apparatus in the cold to minimize the likelihood of exonucleases present on human skin or in airborne bacteria from contaminating the samples and degrading them.³⁷ No precautions, however, can be taken to prevent degradation of the DNA sample recovered from the scene of the crime before it gets to the laboratory. If pronounced, degradation results in loss of all fragments. If less severe, selective loss of particular bands may occur.³⁸ Although it is relatively easy to test for DNA degradation, such a test may not be performed because of the risk that after using up some of the DNA sample in performing the test, an insufficient sample to perform the fingerprinting analysis will remain.

b. Subjective nature of declaring a match

Theoretically, a match should be declared only when the polymorphic fragments in one sample exactly match those in the other. However, gel electrophoresis is insufficiently precise to create an exact match even when the same sample is run twice.³⁹ Furthermore, the bands that indicate the position of fragments are not sharp lines like the bar codes on supermarket items, but are somewhat blurred and curved.⁴⁰ In a research environment, matches are declared by eye. That is, if it appears to the investigator that bands in one track occur in roughly corresponding positions to those in another track, there is a match. Initially, forensic laboratories performing DNA fingerprinting also adopted the eyeball method. This approach was criticized in *People v. Castro*,⁴¹ and more recently, testing laboratories have declared matches by computer analyses of band patterns.⁴² Although the use of computers removes the case-by-case variability of subjective human judgment, matches are still declared on the basis of arbitrary margins of errors defined by the computer's human operators and not on the basis of perfect alignment of bands.

36. See *Castro*, 144 Misc. 2d at 969, 545 N.Y.S.2d at 996. Exonucleases are enzymes that progressively digest the ends of DNA molecules to generate free bases. Exonucleases are found in all living organisms.

37. See generally T. MANIATIS, *supra* note 23.

38. See Lander, *supra* note 33, at 503.

39. Thompson & Ford, *supra* note 23, at 87.

40. *Id.* at 87 n.188.

41. 144 Misc. 2d 956, 967, 545 N.Y.S.2d 985, 995 (1989).

42. See *United States v. Jakobetz*, 747 F. Supp. 250, 259 (D. Vt. 1990).

c. *Statistical analysis*

The overall exclusion frequency is the product of the individual frequencies of occurrence of each matching polymorphic fragment. Application of the product rule assumes, *inter alia*, that the individual frequencies are independent, an assumption that may be invalid. Although individuals do not knowingly choose their marital partners because of their possession of specific polymorphic fragments, humans do not necessarily mate at random with respect to polymorphic fragments.⁴³ The DNA code, of which polymorphic fragments are a part, determines characteristics such as intelligence and race which may have a profound effect on choice of marital partner. Geographic and religious factors can also contribute to a nonrandom assortment of polymorphic fragments. For example, if people from one small town continually intermarry, their descendants will retain distributions of polymorphic fragments similar to the founders of the town, rather than the more random distribution of the total population.⁴⁴ If the assumption of independent distribution of polymorphic fragments is invalid, the statistical calculation will overestimate the exclusion frequency. Such may well have happened in *Texas v. Hicks*⁴⁵ in which the defendant came from a small, inbred Texas town founded by a handful of families.

Even if the statistical calculation of exclusion frequency is reasonably accurate, it may be presented to the jury in a misleading fashion. Astronomical exclusion frequencies, such as one in thirty billion, purportedly derived from DNA fingerprinting, seem extraordinarily high for a complex process in which each step affords a multitude of opportunities for human error.⁴⁶ This is because exclusion frequencies are calculated on the assumption that a match has been correctly declared and do not take into account the possibility of previous human error in declaring the match.⁴⁷ Of course, this can be explained to a jury, but it is a subtle point that might nevertheless be overlooked. There is an additional danger that, as with other forms of statistical evidence, juries will confuse exclusion frequencies with probabilities of guilt, a confusion that may be encouraged by the prosecutor.

43. *Id.* at 260.

44. Thompson & Ford, *supra* note 23, at 86-87.

45. Unpublished case discussed in Lander, *supra* note 33, at 505.

46. See Dodd, *supra* note 5.

47. If matches are declared by computer, the objective margins of error within which the computer is programmed to operate can and should be taken into account in calculating the exclusion frequency. See *infra* text accompanying note 53. However, random human errors (for example, adding the wrong amount of enzyme or cross-contaminating the samples) are not easily taken into account in calculating the exclusion frequency. In practice, the exclusion frequency is calculated on the assumption that such errors did not occur.

3. *Significance of Errors.*—The preceding section lists only a small fraction of the total number of errors that can occur. A more comprehensive account is given by Thompson and Ford.⁴⁸ In light of this vast assortment of potential errors, euphoric pronouncements of the infallibility of DNA fingerprinting are misplaced.⁴⁹ To suggest, however, that use of this technology in court is reminiscent of the Orwellian nightmare of 1984 is an exaggeration.⁵⁰

Although numerous errors can give rise either to extra spurious bands or missing genuine bands, the fact that an unexpected number of bands appeared indicates that some error has likely occurred. Provided that no attempt is made to interpret band patterns containing an anomalous number of bands, errors of this nature can only benefit the defendant in that the evidence will not be used against him.⁵¹ The difficulty arises when someone attempts to decide which bands are anomalous and which are not and attempts to interpret the pattern notwithstanding the obvious anomaly. For example, in *Castro*, the suspect's DNA sample showed two extra bands compared with the sample found at the scene of the crime. If genuine, these bands would have exonerated the suspect. However, the expert interpreting the gel somewhat arbitrarily concluded that these bands were artifacts and declared a match notwithstanding this anomaly. This was one of the reasons that the *Castro* court found the testing laboratory's procedure unacceptable.⁵²

The subjective nature of declaring a match between samples by visual inspection can be reduced by declaring the match by computer scanning. Computers necessarily operate within margins of error programmed by their human operators, but these margins of error can be taken into account in the statistical calculation of the exclusion frequency.⁵³ Thus, looser criteria for declaring a match result in a lower exclusion frequency and less probative evidence.

Although the danger of a nonrandom distribution of polymorphic fragments cannot be ignored, it may not be as pronounced as was once thought.⁵⁴ The FBI, for example, tested subpopulations including Italians, Swedes, Irish, and Amish and found "very small differences" in distributions of polymorphic fragments between them.⁵⁵

48. See Thompson & Ford, *supra* note 23.

49. See, e.g., Note, *Admit It! DNA Fingerprinting Is Reliable*, 26 Hous. L. Rev. 677 (1989).

50. See Note, *The Dark Side*, *supra* note 10, at 465.

51. United States v. Jakobetz, 747 F. Supp. 250, 262 (D. Vt. 1990).

52. People v. Castro, 144 Misc. 2d 956, 967, 545 N.Y.S.2d 985, 997 (1989).

53. *Jakobetz*, 747 F. Supp. at 259.

54. See, e.g., Lander, *supra* note 33, at 504.

55. *Jakobetz*, 747 F. Supp. at 260.

Notwithstanding the above discussion of the reliability of DNA fingerprinting, the ultimate test of accuracy is to perform blind testing of samples of known origin. Surprisingly, only one such analysis has been published. In that test, Lifecodes correctly matched thirty-seven of fifty-one paired samples and called the remaining fourteen pairs inconclusive.⁵⁶ Cellmark correctly matched forty-four of forty-nine pairs, called four pairs inconclusive, and matched one pair erroneously.⁵⁷

B. Techniques Other Than RFLP for Performing DNA Fingerprinting

Although most DNA fingerprinting evidence to date has derived from the RFLP method, a recent state supreme court decision affirmed the admission of such evidence derived from an entirely different technique based on the polymerase chain reaction.⁵⁸ This technique can yield results from smaller samples than can RFLP analysis.⁵⁹ The polymerase chain reaction method shares the same fundamental principle as the RFLP method (*i.e.*, identification is based on detecting differences in polymorphic regions between different individuals), but has little else in common. A further distinct method of performing DNA fingerprinting based on DNA sequencing is being developed.⁶⁰ Although this Note will not discuss these alternative methodologies in detail,⁶¹ they are nevertheless relevant to the interpretation of the Indiana statute. It cannot be assumed that courts which have found RFLP analysis reliable will also find other methods of performing DNA fingerprinting equally reliable, much less that these courts will admit evidence derived from alternative methods without further inquiry.

III. PRINCIPLES AND PROCEDURES FOR DETERMINING ADMISSIBILITY OF SCIENTIFIC EVIDENCE

The substantive reliability of scientific evidence depends on four questions: (1) whether the fundamental principles from which the evidence derives are sound; (2) whether there are reliable procedures for generating

56. Kinoshita, *Misprints: Seeking New Standards for Forensic DNA Typing*, 261 *Sci. Am.* 16 (Aug. 1989).

57. *Id.*

58. Spencer v. Commonwealth, 240 Va. 78, 89, 393 S.E.2d 609, 621, *cert. denied*, 111 S. Ct. 281 (1990).

59. Higuchi, von Beroldingen, Sensabaugh, & Erlich, *DNA Typing from Single Hairs*, 332 *NATURE* 543 (1988).

60. *Id.*

61. These alternatives are discussed at Thompson & Ford, *supra* note 23, at 76-81.

evidence based on those principles; (3) whether the procedure the testing laboratory purports to be following is reliable;⁶² and (4) whether the testing laboratory actually followed the procedure it claims to be following.⁶³ There is a further procedural question as to whom should be entrusted with answering the four substantive questions above: the scientific community, the trial judge, or the jury. The two main tests for determining the admissibility of evidence, the *Frye* test and the relevancy test, take different approaches.

The *Frye* test places the initial responsibility for assessing reliability on the scientific community. Witnesses from the scientific community testify at a pretrial evidentiary hearing to determine admissibility of scientific evidence. The test under *Frye* is whether "the thing from which the deduction is made [has been] sufficiently established to have gained

62. If it is generally accepted that reliable procedures for performing DNA fingerprinting exist, why would a testing laboratory adopt any procedure other than the generally accepted one? Most obviously, because no generally accepted procedure has been spelled out in detail by an authoritative body. When it is said that a generally accepted procedure exists, what is really meant is merely that the component methods of such a procedure, such as restriction digests, hybridization, and Southern blotting, exist and have proved reliable in nonforensic contexts such as research. The task of the testing laboratories is to assemble the component methods into a procedure appropriate to the particular difficulties encountered in forensic DNA analysis. The particular difficulties arise because the samples available for forensic DNA analysis are often small or contaminated. These difficulties are compounded by the need for accuracy in forensic testing, when a defendant's life may rest on the outcome, in contrast to the situation in a research environment, in which a reasonable rate of error is expected and is acceptable.

Some courts tend to merge the third and fourth criteria ((3) and (4) in the text). See *People v. Castro*, 144 Misc. 2d 956, 958-59, 545 N.Y.S.2d 985, 987-88 (1989). This is sensible if a generally accepted procedure is spelled out in detail by an authoritative body or is so obvious that it is self-defining. If this is so, the only remaining question is whether the testing laboratory complied with the generally accepted procedure. However, given the presently abstract nature of what constitutes a generally accepted procedure, the question of whether a laboratory complied with such a procedure is meaningless. It makes more sense to consider as separate questions whether the procedure the testing laboratory purports to be following is reliable and whether the testing laboratory actually followed this purported procedure. Once it is accepted that a particular laboratory's procedure is reliable, courts can focus on the simpler question of whether departures from this accepted procedure are material.

63. Why would a testing laboratory ever depart from the procedure it claims to be following in a particular application of DNA fingerprinting? First, there are many steps in the procedure that offer the opportunity for inadvertent departure from the purported procedure through human error. Second, the procedure involves use of unstable reagents such as enzymes which may not always behave in the expected manner. Third, the testing laboratory might deliberately depart from its usual procedure to attempt to encounter problems arising from an atypical sample, such as a sample that contains an unusual form of contamination.

general acceptance in the particular field in which it belongs.”⁶⁴ The term “thing” is not self-defining; in practice courts have sought to determine whether there is general acceptance of the first, second, and sometimes the third substantive criteria for the evidence in question.⁶⁵ Even if scientific evidence satisfies the *Frye* test as determined by the scientific community, it may still be subject to the traditional evidentiary challenge of legal relevancy as determined by the court.⁶⁶ Evidence is legally relevant if its probative value outweighs its prejudice.⁶⁷ Probative value depends on reliability. Prejudice results when the jury does not attempt to make its own assessment of reliability, but unquestionably accepts scientific evidence because of the expert status of its proponent.⁶⁸ Questions of reliability not resolved by the *Frye* test (*i.e.*, the fourth and possibly the third of the above criteria) can, in principle, be raised to have scientific evidence held legally irrelevant by the court. In balancing probative value against prejudice, much depends on the seriousness of the alleged error and whether it is the kind of error that a jury could take into account in determining what weight to give the scientific evidence. If scientific evidence passes the *Frye* test and is not otherwise held legally irrelevant by the court, its reliability remains open to attack on cross-examination. Such attack goes to the weight of the evidence and the credibility placed on it by the jury.

Court rulings under *Frye* have the same precedential value as other case law. Thus, if a trial court’s *Frye* ruling is affirmed by an appellate court, other trial courts in that jurisdiction are bound by the decision.⁶⁹ This would not, however, preclude trial courts from conducting a *Frye* hearing of more limited scope on different facts, as for example, if testing was performed by a different laboratory than that in the appellate decision. After further appellate decisions, the need for and scope of *Frye* hearings would diminish and eventually disappear.

The *Frye* test is the traditional means of assessing reliability and is said to have the following advantages.⁷⁰ First, it places the primary decision to admit scientific evidence in the hands of those most qualified to make it, that is, the scientific community rather than the court.

64. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923).

65. See *People v. Castro*, 144 Misc. 2d 956, 958, 545 N.Y.S.2d 985, 987 (1989); *People v. Wesley*, 140 Misc. 2d 306, 312, 533 N.Y.S.2d 643, 650 (1988).

66. *Baker v. Wagers*, 472 N.E.2d 218, 219 (Ind. Ct. App. 1984); E. CLEARY, MCCORMICK ON EVIDENCE § 203, at 605 (3d ed. 1984).

67. E. CLEARY, *supra* note 66, § 185, at 545.

68. See Note, *The Frye Doctrine and Relevancy Approach Controversy: An Empirical Evaluation*, 74 GEO. L.J. 1769, 1774 (1986) [hereinafter Note, *The Frye Doctrine*].

69. *State v. Ford*, 301 S.C. 485, 489, 392 S.E.2d 781, 784 (1990).

70. *People v. Kelley*, 17 Cal. 3d 24, 31, 549 P.2d 1240, 1248, 130 Cal. Rptr. 144, 151 (1976).

Second, application of the *Frye* test creates precedent ensuring uniformity of decisions. Third, it is a deliberately conservative test that results in a lag period between the invention of new techniques and the introduction of evidence derived from them into court. The lag period ensures appropriate refinement of new procedures before evidence derived from them becomes admissible.

Although *Frye* is the traditional standard, there is a trend in favor of the alternative relevancy standard.⁷¹ In this approach, the court resolves all reliability questions according to the legal relevancy standard. This approach differs from *Frye* in that the court makes its own appraisal of reliability, rather than attempting to perceive general acceptance in the scientific community. It also differs from *Frye* in that the substantive criteria for determining reliability, which are answered sequentially under *Frye*, tend to merge under the relevancy approach. The court determines reliability by reference to somewhat arbitrary criteria established by prior case law such as the potential rate of errors, the existence and maintenance of standards, the care with which the technique was employed and whether it was susceptible to abuse, analogies with other admissible scientific procedures, and the existence of fail-safe characteristics.⁷² If the court finds that evidence is legally relevant, it is admissible. Reliability can still, of course, be challenged during trial so as to reduce the weight given to the evidence.

The relevancy test results from some jurisdictions' frustration with the inherently conservative nature of the *Frye* test and the accompanying lag period between discovery of a new technique and admissibility of evidence derived from it. Courts applying the relevancy standard are more likely to admit novel scientific evidence than courts applying *Frye*.⁷³

IV. SUMMARY OF COURT DECISIONS ON ADMISSIBILITY OF DNA FINGERPRINTING EVIDENCE

Since 1989, DNA fingerprinting has been the subject of five state supreme court decisions,⁷⁴ five intermediate appellate decisions,⁷⁵ several

71. E. CLEARY, *supra* note 66, § 203, at 606.

72. United States v. Williams, 583 F.2d 1194, 1198 (2d. Cir. 1978), *cert. denied*, 439 U.S. 1117 (1979).

73. See Note, *The Frye Doctrine*, *supra* note 68, at 1771.

74. Caldwell v. State, 260 Ga. 278, 393 S.E.2d 436 (1990); State v. Schwartz, 447 N.W.2d 442 (Minn. 1989); State v. Pennington, 327 N.C. 89, 393 S.E.2d 847 (1990); State v. Ford, 301 S.C. 485, 392 S.E.2d 781 (1990); Spencer v. Commonwealth, 240 Va. 78, 393 S.E.2d 609, *cert. denied*, 111 S. Ct. 281 (1990).

75. Martinez v. State, 549 So. 2d 694 (Fla. Dist. Ct. App. 1990); Andrews v. State, 533 So. 2d 841 (Fla. Dist. Ct. App. 1989); Cobey v. State, 73 Md. App. 233, 533 A.2d 944 (1989); Glover v. State, 787 S.W.2d 544 (Tex. Ct. App. 1990); Kelly v. State, 792 S.W.2d 579 (Tex. Ct. App. 1990).

published New York trial court decisions,⁷⁶ and one published federal district court decision.⁷⁷

The first appellate court to consider DNA fingerprinting held the evidence admissible under the relevancy test.⁷⁸ This decision, however, was somewhat less than satisfactory in that it was based on a trial court record containing testimony from three witnesses in favor of DNA fingerprinting and none opposing it. Furthermore, two of the three witnesses testifying in favor of DNA fingerprinting were employees of the company performing the test and could scarcely be considered disinterested.

The first exacting inquiry into DNA fingerprinting was by the court in *People v. Castro*,⁷⁹ which examined its reliability under *Frye*. After hearing extensive testimony from prosecution and defense expert witnesses, the court found that there was general acceptance in the scientific community of both the theory underlying DNA fingerprinting and the existence of techniques capable of generating reliable results from that theory (prongs one and two of *Frye*).⁸⁰ However, the court held DNA fingerprinting evidence inadmissible (at least for the prosecution, which was seeking to have it admitted) on the grounds that the testing laboratory, Lifecodes, had not properly complied with generally accepted scientific procedures.⁸¹

The Minnesota Supreme Court came to a similar conclusion in *State v. Schwartz*.⁸² This decision was based on an exacting pretrial inquiry at which testimony was heard from twelve expert witnesses. The court found that there is general acceptance of the theory behind DNA fingerprinting and the existence of techniques capable of generating reliable results from that theory.⁸³ As in *Castro*, the *Schwartz* court held the evidence inadmissible because of deficiencies in the procedure of the particular testing laboratory, in this case Cellmark.⁸⁴ The Minnesota Supreme Court also held that statistical evidence from DNA fingerprinting was generally inadmissible, but this was more because of Minnesota's long-standing distrust of statistical evidence than because of specific deficiencies.⁸⁵

76. E.g., *People v. Castro*, 144 Misc. 2d 956, 545 N.Y.S.2d 985 (1989); *People v. Wesley*, 140 Misc. 2d 306, 533 N.Y.S.2d 643 (1988).

77. *United States v. Jakobetz*, 747 F. Supp. 250 (D. Vt. 1990).

78. *Andrews v. State*, 533 So. 2d 841, 850 (Fla. Dist. Ct. App. 1989).

79. 144 Misc. 2d 956, 545 N.Y.S.2d 985 (1989).

80. *Id.* at 979, 545 N.Y.S.2d at 999.

81. *Id.* at 980, 545 N.Y.S.2d at 999.

82. 447 N.W.2d 422 (Minn. 1989).

83. *Id.* at 426.

84. *Id.* at 427.

85. *Id.* at 429. After this decision, the Minnesota legislature made statistical evidence admissible by statute. See MINN. STAT. ANN. § 634.26 (Supp. 1991).

After *Schwartz*, there has been a succession of published opinions approving admission of DNA fingerprinting evidence.⁸⁶ This apparent uniformity following the initial doubts raised in *Castro* and *Schwartz* results from several factors. First, the testing laboratories have improved their procedures in response to the criticisms voiced in *Castro* and *Schwartz*. For example, the FBI has replaced subjective visual inspection with gel-scanning machines that operate according to objective margins of error, and these margins of error are taken into account in calculating the exclusion frequency.⁸⁷ Controls are now included to check for degradation of DNA.⁸⁸ Perhaps most importantly, the FBI now claims to discard all results when the controls indicate a problem, rather than attempting to interpret ambiguous fragment patterns as in *Castro*.⁸⁹

Second, all but one of the recent decisions⁹⁰ have been based on the less-exacting relevance test, rather than the *Frye* test used in *Castro* and *Schwartz*.⁹¹

Third, some of these decisions have been based on trial court records containing testimony from only prosecution expert witnesses, some or all of whom were employees of the testing laboratories.⁹² In other trials, the defendant's so-called experts were woefully inadequate. In *State v. Pennington*,⁹³ the defendant's lone "expert" conceded that he had little time to analyze the testing laboratory's procedure in detail.⁹⁴ In *Kelly v. State*,⁹⁵ the defendant attempted to combat five prosecution witness with a single "expert" whose credentials were confined to a bachelor's

86. *United States v. Jakobetz*, 747 F. Supp. 250 (D.C. Vt. 1990); *Martinez v. State*, 549 So. 2d 694 (Fla. Dist. Ct. App. 1989); *Andrews v. State*, 533 So. 2d 841 (Fla. Dist. Ct. App. 1988); *Caldwell v. State*, 260 Ga. 278, 393 S.E.2d 436 (1990); *Cobey v. State*, 73 Md. App. 944, 533 A.2d 944 (1987); *State v. Pennington*, 327 N.C. 89, 393 S.E.2d 847 (1990); *State v. Ford*, 301 S.C. 485, 392 S.E.2d 781 (1990); *Glover v. State*, 787 S.W.2d 544 (Tex. Ct. App. 1990); *Kelly v. State*, 792 S.W.2d 579 (Tex. Ct. App. 1990); *Spencer v. Commonwealth*, 240 Va. 78, 393 S.E.2d 609, *cert. denied*, 111 S. Ct. 281 (1990).

87. *Jakobetz*, 747 F. Supp. at 259.

88. *Id.* at 257.

89. *Id.*

90. *Ford*, 301 S.C. at 485, 392 S.E.2d at 781.

91. See, e.g., *United States v. Jakobetz*, 747 F. Supp. 250 (D.C. Vt. 1990); *Caldwell v. State*, 260 Ga. 278, 393 S.E.2d 436 (1990); *State v. Pennington*, 327 N.C. 89, 393 S.E.2d 847 (1990); *Kelly v. State*, 792 S.W.2d 579 (Tex. Ct. App. 1990); *Spencer v. Commonwealth*, 240 Va. 78, 393 S.E.2d 609, *cert. denied*, 111 S. Ct. 281 (1990).

92. See, e.g., *Glover v. State*, 787 S.W.2d 544, 548 (Tex. Ct. App. 1990); *Spencer*, 240 Va. at 90, 393 S.E.2d at 620.

93. 327 N.C. 89, 393 S.E.2d 947 (1990).

94. *Id.* at 97, 393 S.E.2d at 853.

95. 792 S.W.2d 579 (Tex. Ct. App. 1990).

degree and a public school teaching certificate.⁹⁶ The same court declined to consider an *amicus* brief challenging the reliability of DNA fingerprinting evidence on grounds that it could not consider error not asserted by the defendant.⁹⁷

Based on *Castro*, *Schwartz*, and *Ford*, there is unanimous agreement that DNA fingerprinting satisfies the first two prongs of the *Frye* test.⁹⁸ The doubts raised by *Castro* and *Schwartz* over the particular procedures of individual testing laboratories have been diminished, but not altogether dispelled by the more recent cases. Although these cases have revealed some improvements in the reliability of testing procedures, the courts' acceptance of DNA fingerprinting evidence also results from inadequate presentation of the defense and the less stringent legal relevancy standard.

V. THE NEW INDIANA DNA FINGERPRINTING STATUTE

A. *Admissibility of DNA Fingerprinting Evidence in Indiana Prior to the Statute*

Indiana uses the *Frye* test for determining the admissibility of novel scientific evidence which means that the relevant scientific community makes the initial determination.⁹⁹ Indiana takes a broad view of what comprises the relevant scientific community.¹⁰⁰ To determine the admissibility of voice spectrograph analysis, the Indiana Supreme Court found that the relevant scientific community is comprised of linguists, psychologists, and engineers, rather than simply technicians who use voice spectrography for identification purposes.¹⁰¹ The adoption of the *Frye* standard applied from the perspective of a broad scientific community results in a particularly conservative approach to the admissibility of novel scientific evidence. In keeping with this approach, the Indiana Supreme Court held that voice spectrographs are inadmissible in Indiana,¹⁰² an issue on which courts in other jurisdictions are split.¹⁰³

There are no appellate decisions on the admissibility of DNA fingerprinting evidence in Indiana. Indiana trial courts confronted with DNA fingerprinting evidence as a matter of first impression will, there-

96. *Id.* at 583.

97. *Id.* at 588.

98. *State v. Schwartz*, 447 N.W.2d 422, 426 (Minn. 1989); *People v. Castro*, 144 Misc. 2d 956, 979, 545 N.Y.S.2d 985, 999 (1989); *State v. Ford*, 301 S.C. 485, 490, 392 S.E.2d 781, 784 (1990).

99. *Cornett v. State*, 450 N.E.2d 498, 503 (Ind. 1983).

100. *Id.*

101. *Id.*

102. *Id.*

103. *Id.* at 502.

fore, be expected to conduct a pretrial *Frye* hearing.¹⁰⁴ The trial court might reduce the scope of a *Frye* hearing by taking notice of nonbinding precedent from other Indiana trial courts or the appellate courts of other jurisdictions. Opinions from other jurisdictions, however, would only be relevant to the extent that they adopted the same criterion for assessing scientific evidence as Indiana, namely the *Frye* test incorporating a broad-based scientific community. This includes *Castro*, *Schwartz*, and possibly *Ford*.¹⁰⁵ The remaining appellate decisions, including those most receptive to DNA fingerprinting, would have little relevance. These decisions were neither based on the *Frye* standard nor on trial court records containing testimony from the broad-based scientific community. Given the criticism of DNA fingerprinting in the cases with strongest precedential value, a *Frye* hearing by an Indiana court would hardly be a formality.

Conducting a full-scale *Frye* hearing each time the admissibility of DNA fingerprinting evidence is sought would be an immense and futile undertaking. (The *Frye* hearing in *Castro* took twelve weeks to complete.) However, time could be saved if Indiana trial courts took notice of nonbinding precedent. Eventually, a decision on admissibility would be appealed to the Indiana Supreme Court. This would establish a uniform standard for the jurisdiction and make clear which, if any, remaining inquiries should be conducted on a case-by-case basis.

B. Enactment of the Statute

The Indiana statute became effective July 1, 1990. There is no published legislative history, nor are there court decisions interpreting the statute. Indiana is the third state to enact a statute governing DNA fingerprinting. The Indiana statute closely follows the wording of a Minnesota statute effective one year earlier.¹⁰⁶ Maryland also has a DNA fingerprinting statute of apparently unrelated genus.¹⁰⁷ New York has

104. See, e.g., State v. Hopkins, No. CCR86-428 (Allen County Ct. 1988) (unpublished opinion cited in Thompson & Ford, *supra* note 23, at 59).

105. State v. Schwartz, 447 N.W.2d 422 (Minn. 1989); People v. Castro, 144 Misc. 2d 956, 545 N.Y.S.2d 985 (1989); State v. Ford, 301 S.C. 485, 392 S.E.2d 781 (1990). The *Ford* court applied the *Frye* test, but it is not clear who testified at the pretrial hearing.

106. In a civil or criminal trial or hearing, the results of DNA analysis, as defined in section 299C.155, are admissible in evidence without antecedent expert testimony that DNA analysis provides a trustworthy and reliable method of identifying characteristics in an individual's genetic material upon a showing that the offered testimony meets the standards for admissibility set forth in the Rules of Evidence.

MINN. STAT. ANN. § 634.25 (Supp. 1991).

107. "In any criminal proceeding, the evidence of a DNA profile is admissible to prove or disprove the identity of any person." MD. CTS. & JUD. PROC. CODE ANN. § 10-915 (Supp. 1990).

an act pending, the effect of which will be to limit the admissibility of DNA fingerprinting to tests performed in state-certified laboratories.¹⁰⁸ Congress is also considering similar legislation for federal courts.¹⁰⁹

C. Possible Interpretations of the Statute

It is obvious that the statute is intended to make DNA fingerprinting evidence more easily admissible in Indiana courts. It is less clear precisely to what extent the statute is intended to restrict the scope of pretrial evidentiary hearings or indeed, whether the statute is intended to eliminate such hearings entirely. The literal wording of the statute read in light of Indiana's common-law approach to admissibility of novel scientific evidence suggests three interpretations.

The statute provides that "the results of forensic DNA analysis [*i.e.*, DNA fingerprinting] are admissible in evidence without antecedent expert testimony that forensic DNA analysis provides a trustworthy and reliable method of identifying characteristics in an individual's genetic material."¹¹⁰ If expert testimony is not required to establish that DNA fingerprinting provides a reliable method, the statute at least reflects a legislative determination that reliable methods for performing DNA fingerprinting can be devised (*i.e.*, that the second prong of the *Frye* test is satisfied). If the legislature finds the second prong of *Frye* satisfied, it would seem that it must also accept the first prong of the *Frye* test (*i.e.*, the fundamental principles underlying DNA fingerprinting are sound). One interpretation of the statute, therefore, represents a legislative determination that DNA fingerprinting satisfies the first two prongs of the *Frye* test. Accordingly, expert testimony might still be required to show that the particular procedure of the testing laboratory is reliable under the third prong of *Frye* and that any departures from this procedure in a particular application are not material as judged by the legal relevancy standard.

There are, however, two other possible interpretations of the statute. The statute might reflect a legislative determination not only that reliable techniques for performing DNA fingerprinting can be devised, but also that the actual procedures of the recognized testing laboratories are reliable. This second interpretation of the statute represents a legislative determination that DNA fingerprinting satisfies all three prongs of the *Frye* test and effectively precludes any further inquiry under *Frye*. Accordingly, expert testimony would only be required to show that any departures from the purported procedure of the testing laboratory in a

108. Slackman, *supra* note 6, at 3.

109. Marcotte, *supra* note 7, at 26.

110. IND. CODE ANN. § 35-37-4-10 (Burns Supp. 1990).

particular application were immaterial as judged by the legal relevancy standard.

A third possible interpretation of the statute is that it represents a legislative determination that antecedent expert testimony should never be required to show the reliability of DNA fingerprinting. Accordingly, DNA fingerprinting evidence would be virtually *per se* admissible.¹¹¹ The reliability of DNA fingerprinting evidence could only be challenged at trial so as to affect the weight given to the evidence, but not its admissibility.

D. Discerning the Intended Meaning of the Statute

1. As a Codification of What Has Been Decided in Other Jurisdictions.—The first interpretation of the Indiana statute, whereby expert testimony is not required to establish the soundness of fundamental principles or the feasibility of devising reliable methods, does no more than codify the unanimous decisions of the courts of other jurisdictions. The courts in *Castro*, *Schwartz*, and *Ford* all found that DNA fingerprinting satisfies the first two prongs of *Frye*.¹¹²

The second interpretation, whereby expert testimony is not required to establish the reliability of the actual procedures of particular testing laboratories, goes somewhat beyond the decisions of other jurisdictions. Courts applying the relevancy standard have held that the procedures of the currently recognized testing laboratories are reliable.¹¹³ However, under the *Frye* test, which is the applicable standard in Indiana, both the Cellmark and Lifecodes procedures have been found unreliable¹¹⁴ (at least where evidence derived from them is sought to be admitted by the prosecution).¹¹⁵ It remains to be seen whether improvements incorporated by Cellmark and Lifecodes since these decisions will lead to acceptance by the same courts that rejected their original procedures.

111. Presumably, admissibility could still be challenged on grounds that do not require refutation by expert testimony. Discrepancies alleged to exist in the chain of custody of DNA samples provide grounds for such a challenge.

112. *State v. Schwartz*, 447 N.W.2d 422, 426 (Minn. 1989); *People v. Castro*, 144 Misc. 2d 956, 979, 545 N.Y.S.2d 985, 999 (1989); *State v. Ford*, 301 S.C. 485, 490, 392 S.E.2d 781, 784 (1990).

113. See, e.g., *United States v. Jakobetz*, 747 F. Supp. 250 (D. Vt. 1990) (FBI); *Caldwell v. State*, 260 Ga. 278, 393 S.E.2d 436 (1990) (Lifecodes); *State v. Pennington*, 327 N.C. 89, 393 S.E.2d 847 (1990) (Cellmark); *Spencer v. Commonwealth*, 240 Va. 78, 393 S.E.2d 609, *cert. denied*, 111 S. Ct. 281 (1990) (Cetus).

114. *Schwartz*, 447 N.W.2d at 427; *Castro*, 144 Misc. 2d at 980, 545 N.Y.S.2d at 999.

115. Although the *Castro* court denied admission of DNA fingerprinting evidence for the prosecution, it indicated that it would admit the same evidence for the defense. *Castro*, 144 Misc. 2d at 979, 545 N.Y.S.2d at 999.

The third interpretation of the statute, whereby DNA fingerprinting evidence is virtually *per se* admissible, finds no support in the opinions of courts of other jurisdictions. In *Ford*, the court found that DNA fingerprinting satisfies the *Frye* test, but stated that “[t]his, however, does not mean that DNA test results should always be admitted into evidence . . . [i]ssues pertaining to relevancy or prejudice may be raised.”¹¹⁶ The *Pennington* court, applying the relevancy test, stated, “While . . . evidence of DNA profile testing is generally admissible and was admissible in the present case, this should not be interpreted to mean that DNA test results should always be admitted into evidence.”¹¹⁷

To summarize, the first interpretation is justified as a codification of case law in other jurisdictions, the second interpretation goes somewhat beyond case law at least where, as has usually been the case,¹¹⁸ admissibility of DNA fingerprinting evidence is sought by the prosecution, and the third interpretation expressly contradicts case law.

2. By Analogy With Other Statutes.—A Minnesota statute on DNA fingerprinting,¹¹⁹ enacted one year prior to the Indiana statute, may aid the interpretation of the Indiana statute. Like Indiana, Minnesota recognizes *Frye* as the standard for determining the admissibility of scientific evidence.¹²⁰ The Indiana statute so closely follows the wording of its Minnesota counterpart to suggest derivation. The Minnesota statute, however, contains two extra clauses that are of considerable assistance in discerning its meaning. First, the Minnesota statute contains a reference to a distinct section of the code which provides that “[t]he bureau shall adopt uniform procedures and protocols to maintain, preserve and analyze human biological specimens for DNA.”¹²¹ If the Minnesota legislature, via “the bureau,” finds it necessary to define its own standards for performing DNA analysis, it can hardly be intending to confer statutory endorsement of the procedures used by the current testing laboratories. Second, the Minnesota statute provides that the “results of DNA analysis . . . are admissible in evidence without antecedent expert testimony that DNA analysis provides a trustworthy and reliable method of identifying characteristics . . . upon a showing that the offered testimony meets the

116. *State v. Ford*, 301 S.C. 485, 490, 392 S.E.2d 781, 784 (1990).

117. *State v. Pennington*, 327 N.C. 89, 102, 393 S.E.2d 842, 854 (1990).

118. There has been only one appellate decision on the admissibility of DNA fingerprinting in which the defendant sought admission. The court refused admission, but this decision was inevitable in view of the testing laboratory’s (Cellmark’s) own report stating that no result could be obtained from its test. *State v. Woodall*, 385 S.E.2d 253 (W. Va. 1989).

119. MINN. STAT. ANN. § 634.25 (Supp. 1991). For the text of this statute, see *supra* note 106.

120. *State v. Schwartz*, 447 N.W.2d 422, 424 (Minn. 1989).

121. MINN. STAT. ANN. § 299C.155 subd. 3 (Supp. 1991).

*standards of admissibility set forth in the Rules of Evidence."*¹²² The extra italicized clause, not present in the Indiana statute, makes it clear that the Minnesota statute is not intended to eliminate traditional evidentiary challenges to DNA fingerprinting.¹²³

If the Minnesota statute is not intended to endorse particular procedures or preclude traditional evidentiary challenges to departures from particular procedures, then it only affects the first two prongs under *Frye*. The Minnesota statute simply represents a legislative determination that DNA fingerprinting meets the first two prongs of *Frye* and consequently, that expert testimony is not required to establish this at each trial.

The similarity in wording between the Indiana and Minnesota statutes suggests that the Indiana legislature may have derived the Indiana statute from the Minnesota version. However, if the Indiana legislature intended its statute to have the same meaning, it is not clear why it did not follow Minnesota in conditioning admissibility on compliance with procedures to be promulgated by an administrative agency (*i.e.*, the "bureau"). Conceivably, this omission might be an oversight. It might also reflect a legislative determination that the particular procedures of testing laboratories should continue to be judged on a common-law basis. In direct contrast to the Minnesota statute, it might reflect a legislative endorsement of the particular procedures of the current testing laboratories. Similarly, if the Indiana legislature intended the same meaning as its Minnesota counterpart, it is not clear why it omitted the clause conditioning admissibility on conformity to the general rules of evidence. This might have been either because the legislature thought it too obvious to mention or because the legislature did not want it to apply.

An earlier Indiana statute regulating the admissibility of evidence derived from blood-group marker tests¹²⁴ may also shed some light on the Indiana DNA fingerprinting statute. Blood-group marker tests yield similar information to DNA fingerprinting, albeit with less sensitivity. The reliability of this technique has also been questioned.¹²⁵ Indiana Code section 31-6-6.1-8 provides that the results of blood-group marker testing are admissible in "all paternity proceedings, *unless the court excludes the results . . . for good cause.*"¹²⁶ This statute has been interpreted to represent a legislative determination that evidence derived

122. MINN. STAT. ANN. § 634.25 (Supp. 1991) (emphasis added).

123. *See supra* text accompanying note 66.

124. IND. CODE ANN. § 31-6-6.1-8 (Burns 1989).

125. *See* Thompson & Ford, *supra* note 23, at 47.

126. IND. CODE ANN. § 31-6-6.1-8 (Burns 1989) (emphasis added).

from blood-group marker testing satisfies the *Frye* test,¹²⁷ but that the reliability of the specific test in question must still be demonstrated before its results are admissible in evidence.¹²⁸ One might expect the legislature to treat the admissibility of similar technologies (*i.e.*, DNA fingerprinting and blood-group marker analysis) in the same manner. Accordingly, the DNA fingerprinting statute would reflect a legislative determination that DNA fingerprinting satisfies the *Frye* test (or at least the first two prongs thereof), but that it remains open to traditional evidentiary challenge. Yet, unlike the blood-marker statute, the DNA fingerprinting statute does not contain a clause conditioning admissibility on the discretion of the court. As in the discussion of the Minnesota statute, the question arises whether the legislature omitted such a clause because it was too obvious to mention or because it was not to apply.

3. *On Policy Grounds.*—It is obvious that the purpose of the statute is to facilitate the admission of DNA fingerprinting evidence into Indiana courts by reducing the expert testimony required to establish a foundation of reliability. Two policies might justify this purpose. One policy is to save the time and expense of pretrial evidentiary hearings to determine admissibility.¹²⁹ The pretrial hearing in *Castro*, for example, took twelve weeks to complete, and the hearing in *Schwartz* required testimony from twelve different expert witnesses.¹³⁰ A second policy is to ensure that DNA fingerprinting evidence actually gets before the court and is not waylaid by spurious arguments about reliability in a pretrial evidentiary hearing.

The first interpretation of the statute, whereby DNA fingerprinting satisfies the first two prongs of the *Frye* test, represents a codification of what has been unanimously decided by the courts of other jurisdictions. Accordingly, the statute might save some of the time and expense of the pretrial *Frye* hearings that would have to be held until such time as the Indiana Supreme Court announced its seemingly inevitable agreement with other jurisdictions. Yet, if the statute merely forecloses debate of the first two prongs of the *Frye* test, the time savings will not be great. In previous pretrial hearings on DNA fingerprinting, argument has centered on the particular procedure of the testing laboratory and

127. The court did not explicitly state which prongs of the *Frye* test were satisfied. The distinction is more important for more complicated procedures (such as DNA fingerprinting) that are subject to many possible variations.

128. *Baker v. Wagers*, 472 N.E.2d 218, 219 n.1 (Ind. Ct. App. 1984).

129. The expense of a protracted pretrial hearing may be a more significant concern to a defendant than to the state. This may explain why, in the great majority of cases, DNA fingerprinting evidence has been offered by the state.

130. *State v. Schwartz*, 447 N.W.2d 422, 425 (Minn. 1989); *People v. Castro*, 144 Misc. 2d 956, 957, 545 N.Y.S.2d 845, 986 (1989).

not on underlying theory.¹³¹ Furthermore, although expert witnesses may be able to testify about a particular testing procedure without mentioning general principles, a trial judge is in no position to understand such testimony without this background. Thus, in practice, a legislative determination that DNA fingerprinting satisfies the first two prongs of the *Frye* test will result in only limited savings of time. The issues on which argument is foreclosed are noncontentious and the information concerning these issues must still be presented to the court in some form for the court to understand argument on the contentious issues.

If the first interpretation of the statute will result in little saving of court time, it will have still less effect in reducing the opportunity for preventing admissibility on the basis of spurious challenges to reliability. Such challenges are likely to be directed at the contentious issues, on which, in this interpretation of the statute, argument is not foreclosed.

To save significant court time and to appreciably reduce the opportunity for spurious objections, argument on the reliability of particular procedures must be foreclosed. The second interpretation of the statute has precisely this effect. If it is accepted that the purported procedure of a testing laboratory is valid, the only remaining question is whether the testing laboratory actually performed the procedure as specified. For major departures from the purported procedure, the trial court could rule the evidence legally irrelevant at a pretrial hearing. Minor departures from purported procedure could be argued to the jury. In either event, the issue is not a *de novo* evaluation of the reliability of the entire procedure, but is the more limited question of whether a departure from that procedure was significant. The saving of time and reduction of spurious argument that will result from this interpretation will be significant, but not without cost.

The second interpretation of the statute will cause considerable confusion as to which laboratory procedures the legislature intended to endorse. The statute does not describe a mechanism for promulgating procedures (in contrast to the Minnesota statute), and it does not name authorized laboratories. At present, there are three laboratories performing DNA fingerprinting by the RFLP method and one by a totally different method. Even between the three laboratories performing RFLP analysis, there are differences in detail such as choice of probe and quality control procedures. In the future, it is likely that further procedures for performing DNA fingerprinting will be devised (*e.g.*, based on DNA sequencing) and that the existing technologies will be licenced to other laboratories.¹³² In this situation, a collective endorsement of DNA fingerprinting procedures is meaningless.

131. See, e.g., *Schwartz*, 447 N.W.2d at 426.

132. Note, *The Dark Side*, *supra* note 10, at 500.

A further cost of foreclosing objections to the reliability of particular procedures is the risk that such objections are not spurious. At least among courts applying the *Frye* test, it is premature to say that a consensus has been reached as to the reliability of particular procedures. If there are remaining doubts about procedures, continued judicial scrutiny may serve to encourage the testing laboratories to improve their procedures and eliminate these doubts. Ironically, the statute does not apply to civil applications of DNA fingerprinting where concerns over reliability are much diminished.¹³³

The costs imposed by the second interpretation of the statute could be somewhat reduced by alternative drafting. The legislature could eliminate confusion as to which laboratories have acceptable procedures by expressly naming those laboratories and periodically updating the list. Alternatively, the legislature could follow its Minnesota counterpart's example of instructing an administrative agency to promulgate its own procedures and regulations. Of course, instructing an administrative agency to devise procedures does not solve the problem of what these procedures should be.

The third interpretation of the statute, whereby DNA fingerprinting evidence is virtually *per se* admissible, would save even more court time than the second and would virtually eliminate challenges to reliability. The difference between the third and second interpretations is that in the former, the reliability of a particular application of a testing procedure will always be left to the jury regardless of alleged departures from purported procedure. In the second interpretation, if the departure is sufficiently egregious, the trial court could hold the evidence legally irrelevant. Eliminating traditional evidentiary challenges to DNA fingerprinting saves time and money and increases predictability. The policy question is whether the saving of time and increased predictability of a *per se* rule outweigh the possibility that a jury will be unable to take into account even egregious departures from purported procedure and the consequent prejudice to the side opposing admissibility.

DNA fingerprinting is a complex process that is not easily understood by those untrained in molecular genetics. Perhaps with the aid of colorful diagrams and analogies, an expert witness can impart some understanding to a lay jury of the fundamental principles.¹³⁴ Yet, to attempt to argue that while the overall principles are sound, departures from purported procedure can make the evidence unreliable, is more difficult. For example, an argument that too much or too little enzyme was added and

133. See *supra* note 8. Note that MINN. STAT. ANN. § 634.25 (Supp. 1991) applies to both civil and criminal trials.

134. Note, *The Dark Side*, *supra* note 10, at 512.

of its possible consequences to the fragment pattern, might appear as nit-picking to a jury who may get lost in a "mire of details and confusion."¹³⁵ There is a considerable risk that the jury will be so overwhelmed by the credentials of the expert presenting the DNA fingerprinting evidence that it will be oblivious to even egregious errors that opposing counsel attempts to draw to its attention. This possibility suggests that at least when the prosecution seeks admissibility of DNA fingerprinting evidence, the defendant should be allowed to challenge departures from a purported procedure in a pretrial hearing. Furthermore, because of the complexity of the DNA fingerprinting procedure, the court should require a stronger rebuttal of alleged departures than would be necessary for other forensic procedures which are less likely to confuse the jury.¹³⁶

Arguably, there is less risk of prejudice when DNA fingerprinting evidence is sought to be admitted by the defendant. In part, this is because society is more concerned with convicting an innocent defendant than with freeing a guilty one. There is, however, a further distinction. When DNA fingerprinting evidence is sought to be admitted by the defendant, the prosecution must have independent incriminating evidence to justify an indictment. This evidence may help to dispel any notion that DNA fingerprinting is infallible. When DNA fingerprinting evidence is sought to be admitted by the prosecution, it may be the only significant evidence in the case, and a jury will be more likely to accept its infallibility.

E. Constitutional Validity of the Statute

Insofar as the statute may facilitate the conviction of a defendant on the basis of unreliable evidence, it is subject to constitutional challenge. The susceptibility to challenge depends on the interpretation of the statute and is most acute for the third interpretation, whereby DNA fingerprinting evidence is virtually *per se* admissible.

First, the statute may be challenged as a violation of the confrontation clause.¹³⁷ It is difficult to effectively cross-examine an expert witness on the reliability of DNA fingerprinting if she is allowed to state the results of a test without antecedent testimony of reliability. This difficulty is exacerbated by commercial laboratories' reluctance to allow pretrial discovery for proprietary reasons.¹³⁸ In these circumstances, it could be argued that an expert witness is not available for cross-examination as

135. *Id.* at 513.

136. *United States v. Jakobetz*, 747 F. Supp. 250, 262 (D. Vt. 1990).

137. "In all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against him. . . ." U.S. CONST. amend. VI.

138. *State v. Schwartz*, 447 N.W.2d 422, 428 (Minn. 1989).

required by the confrontation clause. However, the Supreme Court has held that even a witness who has no memory of events in an out-of-court statement is available for cross-examination.¹³⁹ There is apparently no requirement that a witness supply an antecedent factual context to facilitate cross-examination. It would seem that an expert witness with a complete butunnarrated source of knowledge is more available for cross-examination than a witness with no memory of the underlying events. Thus, a challenge based on the confrontation clause will probably fail.

The statute may also be challenged as a violation of due process. It could be argued that when reliability of evidence is questionable, the defendant has a due process right to a pretrial hearing to determine legal relevancy of the evidence outside the presence of the jury. In some circumstances, the Supreme Court has recognized that when reliability of evidence is in doubt, the defendant has a due process right to a pretrial hearing. In *Jackson v. Denno*,¹⁴⁰ the Supreme Court held that when a genuine issue of voluntariness of a confession is raised, due process requires that the defendant be granted a hearing out of the presence of the jury to determine voluntariness and hence, admissibility.¹⁴¹ The Court noted that involuntary confessions are inadmissible both because they are likely to be unreliable and because of the "strongly felt attitude . . . that important human values are sacrificed where an agency of the government, in the course of securing a conviction, wrings a confession out of an accused against his will."¹⁴²

The due process right to a pretrial hearing was, however, limited in *Watkins v. Sowders*.¹⁴³ The *Watkins* Court held that no due process right to a pretrial hearing existed when the defendant alleged that identification evidence was tainted by suggestive out-of-court identification procedures.¹⁴⁴ The Court distinguished *Jackson* on the grounds that an involuntary confession was not excluded simply because it was unreliable, but because of society's abhorrence of the manner of its extraction.¹⁴⁵ The Court found that infirmity of suggestive identification procedures only affected the reliability of the evidence which could adequately be drawn to the jury's attention by cross-examination.¹⁴⁶

The effect of the Indiana statute is to reduce or eliminate a defendant's right to a pretrial hearing and thereby arguably allow legally

139. *United States v. Owens*, 484 U.S. 555, 564 (1988).

140. 378 U.S. 368 (1964).

141. *Id.* at 377.

142. *Id.* at 386 (quoting *Blackburn v. Alabama*, 361 U.S. 199, 206-07 (1960)).

143. 449 U.S. 341 (1981).

144. *Id.* at 349.

145. *Id.* at 347.

146. *Id.* at 348.

irrelevant evidence to go to the jury subject only to doubts cast by the defendant in cross-examination. As in *Watkins*, the alleged infirmity goes only to reliability, suggesting that there is no violation of due process. However, *Watkins* may be distinguishable. Although a jury may be capable of understanding how suggestive procedures can taint identification evidence, it may not be capable of appreciating how nuances of molecular biology and population genetics can affect the reliability of DNA fingerprinting evidence.

F. Impact of the Statute on Future Trials

The impact of the statute on future trials depends on how the courts interpret it. The first interpretation of the statute, whereby it represents a legislative determination that DNA fingerprinting satisfies the first two prongs of the *Frye* test, may result in a small saving of trial time by abbreviating testimony required for noncontentious issues. The more contentious issue of the reliability of particular procedures will continue to be resolved on a common-law basis. Such an approach may waste court time and may, in some instances, lead to the rejection of DNA fingerprinting evidence based on spurious arguments. However, continued judicial scrutiny should provide an incentive for testing laboratories to further improve the reliability of their procedures. Eventually, trial court decisions on the reliability of particular procedures will be appealed to higher courts creating precedent, and the need for and scope of pretrial hearings will diminish.

The second interpretation of the statute, whereby it represents a legislative endorsement of particular testing procedures, will save appreciable court time and reduce the opportunity for spurious argument. However, absent consensus among the courts of other jurisdictions following *Frye* that particular procedures are reliable, such endorsement may be premature. Its effect may be to exclude spurious and well-founded objections alike. Furthermore, present legislative endorsement of particular procedures removes an incentive for future improvements in reliability. The second interpretation will also create confusion as to which laboratory procedures the legislature intended to endorse. This confusion will increase as the technology is licensed to other laboratories and other methods for performing DNA fingerprinting are devised.

The third interpretation of the statute, whereby DNA fingerprinting evidence is virtually *per se* admissible, will probably be challenged as a violation of due process. This interpretation of the statute is likely to survive constitutional challenge, but it has little else to commend it. DNA fingerprinting evidence is likely to influence the jury, notwithstanding defects sought to have been revealed by cross-examination, with the result that defendants may be convicted on the basis of evidence that should have been excluded as legally irrelevant.

VI. CONCLUSION

The Indiana DNA fingerprinting statute is poorly drafted. Although the statute's purpose of facilitating the admission of DNA fingerprinting evidence is obvious, the extent to which it reduces the expert testimony required to establish a foundation of reliability is less clear. Case law from other jurisdictions and analogous statutes do not unambiguously reveal which of three possible interpretations the Indiana legislature intended. Policy concerns suggest that the statute should be given the most limited of these interpretations. Accordingly, it reflects a legislative determination that the principles underlying DNA fingerprinting are sound and that reliable procedures for performing DNA fingerprinting can be devised. The reliability of particular procedures of specific testing laboratories should continue to be judged on a common-law basis. This interpretation of the statute will result in some saving of court time by reducing the need for testimony on noncontentious issues. More court time could be saved, albeit at the possible risk of eliminating well-founded challenges to reliability, through statutory endorsement of particular procedures. The present statute could, but should not, be interpreted to be such an endorsement. If the legislature wishes to endorse particular procedures, it should at least name the laboratories performing approved procedures, or preferably, promulgate its own approved procedures via an administrative agency. Regardless of whether the present statute is interpreted as an endorsement of particular procedures, it should not be regarded as a *per se* rule of admissibility. The possibility of traditional evidentiary challenge to individual applications of particular procedures should be left open.

JOE LIEBESCHUETZ

An Analysis of *Koske v. Townsend Engineering*: The Relationship Between the Open and Obvious Danger Rule and the Consumer Expectation Test

INTRODUCTION

On December 28, 1979, Margaret Koske severely injured her hand while operating a meat Skinner/slasher machine designed and manufactured by Townsend Engineering.¹ At the time of her accident, Margaret had been employed by the Wilson Foods Company, a meat-packing plant, for six years. Her primary responsibility was trimming excess hair, skin, and abscesses off the jowls after they exited the Skinner/slasher machine. In addition, she assisted on the machine about twice a week because it regularly jammed, causing a bottleneck in production. On the day of her injury, Margaret was assisting to alleviate such a backlog. The jowls were stiff from hanging in the freezer. As a result, they would not automatically feed into the machine. An external force was required for the blades to engage the jowls and pull them into the machine. Because Margaret knew her hands should not be close to the blades, she used one jowl to push another into the machine. The jowl she used to push the others became wet from the conveyor belt. It slid over the top of the jowl closest to the blades, and Margaret's hand became caught in the machine.

Margaret filed suit against Townsend Engineering, alleging strict liability in tort for a design defect pursuant to the 1978 Indiana Product Liability Act.² The trial court granted summary judgment in favor of the defendant on the ground that Margaret's product liability action was barred by the open and obvious danger rule. In affirming the trial court decision, the Indiana Court of Appeals held: (1) that the Indiana Product Liability Act incorporates the open and obvious danger doctrine and (2) that the open and obvious danger doctrine applies to design defect cases.

1. The meat slicing machine cut the skin from pork jowls while slashing the top of the jowl to reveal hidden abscesses. The machine was waist high with a two foot conveyor belt extending from the front. It had 17 circular blades that slashed across the top of the jowl and one long blade that skinned the bottom. The conveyor moved the jowl into the blades. Notches in the blades engaged the jowl and forced it through the machine. A cover through which the operator could see the blades was mounted over the top of the blades. However, no safety mechanism such as a hand guard or a deactivation button was provided at the point of operation.

2. IND. CODE §§ 33-1-1.5-1 to -8 (1988). Margaret also alleged willful and wanton misconduct by the manufacturer for failure to issue post-sale warnings or to recall the machine in reckless disregard of known probable consequences.

In addition, the Indiana Court of Appeals stated that whether a danger is open and obvious is, at times, a question of fact. However, the court also held that when a genuine issue of fact is not presented, the question becomes a matter of law. On transfer, the Indiana Supreme Court reversed the entry of summary judgment and remanded the case for further proceedings, holding that the open and obvious danger rule does not apply to strict liability claims under the Indiana Product Liability Act.³

This Note examines the soundness and potential impact of the decision rendered by the Indiana Supreme Court in *Koske v. Townsend Engineering*.⁴ The focus is on the relationship between the open and obvious danger rule and the consumer expectation test embodied in the Indiana Product Liability Act.⁵ Both concepts employ an objective standard, applied from the consumer's point of view, to determine whether a product is defective. Apparently frustrated by the harsh outcome of the open and obvious danger rule, the Indiana Supreme Court declared the rule inapplicable to strict product liability cases.⁶ The court did so, however, without analyzing the rule's close connection to the consumer expectation test. Although this Note recognizes problems with both the open and obvious danger rule and the consumer expectation test, its main purpose is to illustrate the interrelatedness of the two doctrines. Because this close relationship exists and because the Indiana legislature has embodied a consumer expectation test in the Product Liability Act, the *Koske* court erred when it rejected the open and obvious danger rule.

Section I provides a history of the consumer expectation test. Section II defines the "open and obvious danger rule" and recounts its roots in the area of strict liability. Section III describes the interrelatedness of the open and obvious danger and the consumer expectation tests. Section IV discusses the specific holdings of the Indiana Supreme Court in *Koske*. Section V analyzes the court's rationale for these holdings and pinpoints problems with the decision. Finally, this Note concludes with possible solutions to the dilemma surrounding the open and obvious danger rule.

I. THE CONSUMER EXPECTATION TEST

A. *Indiana Common Law*

Before looking at the history of the open and obvious danger rule, it is necessary to explore the standard used in Indiana to determine

3. *Koske v. Townsend Eng'g*, 551 N.E.2d 437, 442 (Ind. 1990).

4. 551 N.E.2d 437 (Ind. 1990).

5. The consumer expectation test is the standard set forth in the Act for determining whether a product is in a defective condition unreasonably dangerous. IND. CODE § 33-1-1.5-2.5 (1988).

6. *Koske*, 551 N.E.2d at 442.

whether a product is defective. This exploration should begin with the Second Restatement of Torts section 402A. In 1973, the Indiana Supreme Court adopted section 402A as the law regarding products liability.⁷ Section 402A states:

- (1) One who sells any product in a defective condition unreasonably dangerous to the user or consumer or to his property is subject to liability for physical harm thereby caused to the ultimate user or consumer, or to his property, if
 - (a) the seller is engaged in the business of selling such a product, and
 - (b) it is expected to and does reach the user or consumer without substantial change in the condition in which it is sold.
- (2) The rule stated in Subsection (1) applies although
 - (a) the seller has exercised all possible care in the preparation and sale of his product, and
 - (b) the user or consumer has not bought the product from or entered into any contractual relation with the seller.⁸

The rule is intended to apply "only where the product is, at the time it leaves the seller's hands, in a condition not contemplated by the ultimate consumer, which will be unreasonably dangerous to him."⁹ In addition, the defective condition of the product must be unreasonably dangerous to the consumer.¹⁰ Unreasonably dangerous is defined as "dangerous to an extent beyond that which would be contemplated by the *ordinary consumer* who purchases it, with the ordinary knowledge common to the community as to its characteristics."¹¹ Comment i to section 402A cites good whiskey as an example. Whiskey is not unreasonably dangerous merely because it makes some people drunk and it is especially dangerous to alcoholics; however, whiskey containing a dangerous amount of fusel oil is unreasonably dangerous.¹²

Both comments g and i focus on the "consumer's expectations" for the product. The language of comment i denotes that the consumer expectation test is an objective standard. The phrase "contemplated by the *ordinary consumer* who purchases it, with the *ordinary knowledge* common to the community as to its characteristics"¹³ indicates that the

7. Ayr-Way Stores, Inc. v. Chitwood, 261 Ind. 86, 300 N.E.2d 335 (1973).

8. RESTATEMENT (SECOND) OF TORTS § 402A (1965).

9. *Id.* § 402A comment g.

10. *Id.* § 402A comment i.

11. *Id.* (emphasis added).

12. *Id.*

13. *Id.* (emphasis added).

inquiry is what the average consumer would contemplate, not the subjective appreciation of the particular plaintiff alleging the defect.

B. Criticism of the Consumer Expectation Test

The policy underlying the consumer expectation test is that the seller should not become an insurer of his products with respect to all harm generated by their use.¹⁴ The test, however, has been criticized. First, it gives the impression that a product must be specially or unusually dangerous.¹⁵ Second, it is not fully suitable in situations where the consumer does not know what to expect because he does not know the product could be made more safe.¹⁶ Finally, the consumer expectation test has been criticized for imposing a negligence standard on a strict liability statute.¹⁷

C. Indiana Statutory Law

Despite these criticisms, the Indiana legislature adopted the consumer expectation test when it enacted the Product Liability Act in 1978.¹⁸ The statute adopted the Second Restatement of Torts section 402A nearly word for word.¹⁹ In defining “unreasonably dangerous,” the statute refers to “any situation in which the use of a product exposes the user or consumer to a risk of physical harm to an extent beyond that contemplated by the *ordinary consumer* who purchases it with the *ordinary knowledge* about the product’s characteristics common to the community of consumers.”²⁰ Section 2.5 of the statute clearly sets forth an objective standard:

14. R. CARTWRIGHT & J. PHILLIPS, PRODUCTS LIABILITY § 5.16 (1986).

15. J. BEASLEY, PRODUCTS LIABILITY AND THE UNREASONABLY DANGEROUS REQUIREMENT 84 (1981). The way § 402A defines “defective condition” and “unreasonably dangerous” could lead one to believe that a product must be defective *and* unreasonably dangerous.

16. *Id.* This is especially true in design cases and in cases involving complex or novel products. A layperson may not have the capacity to understand the function of the product.

17. Strict liability was designed to relieve the plaintiff from problems of proof inherent in pursuing a negligence theory in products liability cases. Imposing on the plaintiff the burden of proving that the product was (1) defective *and* (2) unreasonably dangerous increases his burden and defeats the purpose of strict liability. R. CARTWRIGHT & J. PHILLIPS, *supra* note 14, at 510-12.

18. IND. CODE §§ 33-1-1.5-1 to -8 (1988).

19. “The only significant departure from the language of section 402A was that section 3(a) of the 1978 Act added the phrase ‘if the user or consumer is in the class of persons that the seller should reasonably foresee as being subject to the harm caused by the defective condition,’ which phrase is not contained in Section 402A.” Koske v. Townsend Eng’g, 551 N.E.2d 437, 442 n.1 (Ind. 1990).

20. IND. CODE § 33-1-1.5-2 (1988) (emphasis added).

(a) A product is in a defective condition under this chapter if, at the time it is conveyed by the seller to another party, it is in a condition:

- (1) not contemplated by *reasonable persons* among those considered expected users or consumers of the product; and
- (2) that will be unreasonably dangerous to the expected user or consumer when used in reasonably expectable ways of handling or consumption.²¹

The effect of the statute is to make section 402A the law in Indiana. In enacting the 1978 statute, the legislature declared that it was codifying and restating the common law of Indiana.²² When making such a declaration, the legislature is presumed to know the common law and to have intended to carry it into the statute except when it expressly indicates otherwise.²³

To be actionable under section 402A, the injury-producing product must be unreasonably dangerous.²⁴ "Unreasonably dangerous" was interpreted at common law to mean dangerous to an extent beyond that which would be contemplated by the ordinary consumer.²⁵ After the passage of the 1978 Indiana Products Liability Act, the common-law interpretation was used to define "unreasonably dangerous."²⁶ Furthermore, this phrase has been defined in terms of an objective standard. The requirement that the product be unreasonably dangerous focuses on the reasonable contemplation and expectations of the ordinary consumer.²⁷ Having surveyed the consumer expectation test, as adopted in Indiana by statute, the next step is to analyze the open and obvious danger rule.

II. OPEN AND OBVIOUS DANGER RULE

As formulated by the Indiana Supreme Court in *Bemis Co. v. Rubush*,²⁸ the open and obvious danger rule provides:

21. *Id.* § 33-1-1.5-2.5(a) (emphasis added).

22. *Masterman v. Veldman's Equip., Inc.*, 530 N.E.2d 312, 315 (Ind. Ct. App. 1988).

23. This rule is a corollary to the rule that statutes in derogation of the common law are to be strictly construed. *State Farm v. Structo Div., King Seeley Thermos Co.*, 540 N.E.2d 597, 598 (Ind. 1989).

24. *Bemis Co. v. Rubush*, 427 N.E.2d 1058, 1061 (Ind. 1981), *cert. denied*, 459 U.S. 825 (1982).

25. *Id.*

26. See, e.g., *Corbin v. Coleco Indus.*, 748 F.2d 411 (7th Cir. 1984) (common-law interpretation used to define the statutory use of "unreasonably dangerous"). The 1983 amendments to the Indiana Product Liability Act included a definition of "unreasonably dangerous."

27. *Jarrell v. Monsanto Co.*, 528 N.E.2d 1158, 1167 (Ind. Ct. App. 1988).

28. 427 N.E.2d 1058, 1061 (Ind. 1981).

In the area of products liability, based upon negligence or based upon strict liability under § 402A of Restatement (Second) of Torts to impress liability upon manufacturers, the defect must be hidden and not normally observable, constituting a latent danger in the use of the product. Although the manufacturer who has actual or constructive knowledge of an unobservable defect or danger is subject to liability for failure to warn of the danger, he has no duty to warn if the danger is open and obvious to all.²⁹

In *Bemis*, the plaintiff, Gerald Rubush, worked as a bagger on a fiberglass insulation batt packing machine. The machine was designed by the Bemis Company. On October 19, 1971, while working as a bagger on the machine, Rubush was struck in the head by the shroud, a visible, moving part of the machine. Rubush sustained serious injuries to his skull and brain.

Neither Rubush nor his co-workers were able to explain exactly what happened. No evidence existed to indicate that the machine malfunctioned or that it was defective in its operation. Similar machines were used for ten years prior to the accident without incident. After the accident, the machine was tested for malfunctions by various experts. The testing, however, revealed no mechanical or electrical malfunctions.

Rubush alleged that the machine was dangerous because it allowed the shroud to descend while any object or person was within its path of operation. He argued that Bemis's failure to design the machine to stop the shroud from descending if something was in its path constituted a design defect. Rubush admitted, however, that the descending shroud was an open and obvious danger.

Bemis contended that it was not strictly liable under section 402A because any dangers of the packing machine were open and obvious. Bemis presented evidence that clearly indicated that the danger was patent. In addition, Bemis's evidence established that protective devices were not feasible and that an alternate design with more remote control buttons would have made the machine more dangerous.

Relying on the Second Restatement of Torts section 402A as the basis for strict liability, the *Bemis* court held that "to be actionable under § 402A, the injury-producing product must be unreasonably dangerous, that is, dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it, with the ordinary knowledge common to the community as to its characteristics."³⁰ This test focuses the inquiry on the expectations of the consumer. Thus,

29. *Id.*

30. *Id.* (emphasis added).

the *Bemis* court's test for determining a "defective condition unreasonably dangerous" is the consumer expectation test.

Next, the *Bemis* court found support for the open and obvious danger rule in Indiana cases and federal cases applying Indiana law.³¹ "[T]o impress liability upon manufacturers, the defect must be hidden and not normally observable, constituting a latent danger in the use of the product."³² The court's rationale was that there must be reasonable freedom and protection for the manufacturer because he is not an insurer against accidents and is not obligated to produce only accident-proof machines.³³ Instead, the manufacturer's duty is to avoid hidden defects or dangers.³⁴

Later cases clarified the objective nature of the open and obvious rule. The objective test, based upon what the user should have known, is used to determine whether a defect or danger is open and obvious.³⁵ Although in many cases this question is a matter of law, this is not absolute;³⁶ sometimes the determination is a question for the trier of fact.³⁷

Courts are split on the proper application of the open and obvious danger rule. "To impress liability upon manufacturers, the defect must be hidden and not normally observable."³⁸ This indicates that one element of the plaintiff's *prima facie* case is to show that the defect is latent.

31. *Id.* See, e.g., *Burton v. L.O. Smith Foundry Prods. Co.*, 529 F.2d 108 (7th Cir. 1976) (when danger or potential danger is known or should be known to the user, there is no duty to warn); *Posey v. Clark Equip. Co.*, 409 F.2d 560 (7th Cir. 1969) (no duty to warn because those receiving warning would normally realize the danger without the warning); *J.I. Case Co. v. Sanderfur*, 245 Ind. 213, 197 N.E.2d 519 (1964) (manufacturer had a common-law duty to protect third parties using the product against hidden defects and dangers).

32. *Bemis*, 427 N.E.2d at 1061.

33. *Id.* at 1062. Note that the policy underlying the open and obvious danger rule is essentially the same as the consumer expectation test.

34. *Id.*

35. *Ragsdale v. K-Mart Corp.*, 468 N.E.2d 524, 527 (Ind. Ct. App. 1984). The fact that a mower blade is not clearly exposed to the user does not make it a latent danger. The ordinary user of a lawn mower is aware of the presence of a blade under the hood of the mower which moves to cut the grass. The court held that such a blade poses an open and obvious danger as a matter of law to one placing a hand into the running mower.

36. *Bridgewater v. Economy Eng'g*, 486 N.E.2d 484, 488 (Ind. 1985).

37. *Id.* Clarifying its holding in *Bemis*, the Indiana Supreme Court cited to *Hoffman v. E.W. Bliss Co.*, 448 N.E.2d 277 (Ind. 1983), as an example of when the question is for the trier of fact. Although the operator of a press would know of the danger of putting one's hand in the press, it was not open and obvious as a matter of law that an internal dysfunction of the press might cause it to activate and recycle itself when it did not do so during normal operation.

38. *Bemis Co. v. Rubush*, 427 N.E.2d 1058, 1061 (Ind. 1981).

If the plaintiff cannot demonstrate this, then the plaintiff fails to meet his burden of production. Nevertheless, many courts and commentators misapply the rule by calling it an affirmative defense.³⁹

III. INTERRELATEDNESS OF THE OPEN & OBVIOUS DANGER RULE AND THE CONSUMER EXPECTATION TEST

The policy rationale supporting the open and obvious danger rule is similar to that advanced by the consumer expectation test. First, a manufacturer is not an insurer against accidents.⁴⁰ Second, when a danger is obvious, a manufacturer can reasonably expect users to act to avoid injury.⁴¹ This rationale is particularly appropriate in warning cases. When a danger is fully obvious and generally appreciated, nothing of value is added by a warning.⁴² One commentator has even postulated that the rule reduces cost and error because it functions as a workable and reliable surrogate for the assumption of risk defense, which because of the defense's intimate connection with subjective states, is difficult to establish by reliable evidence.⁴³

Like the consumer expectation test, the open and obvious danger rule has received criticism from many courts. One court denounced the rule for making "obviousness" the sole determinant of the reasonableness of a danger, rather than one of many factors.⁴⁴ In addition, the product becomes insulated from liability simply because it is patently dangerous.⁴⁵ A victim could never recover for harm suffered as a result of a design hazard that was open and obvious. Consequently, patently dangerous products may be deemed nondefective despite the fact that a safer design was available at only a slight increase in manufacturing costs.⁴⁶ Lastly, in many situations, especially those involving design matters, the consumer would not have safety expectations because he would have no idea how safely the product could be made.⁴⁷ Despite the negative treatment of

39. See *Rogers v. R.J. Reynolds Tobacco Co.*, 557 N.E.2d 1045, 1052 (Ind. Ct. App. 1990) ("The 'open and obvious' defense of *Bemis v. Rubush* . . . has no application to this case for two reasons."); R. EPSTEIN, MODERN PRODUCTS LIABILITY LAW 145-46 (1980) ("The traditional cases accepted the absolute status of the hard-edged open and obvious defense.").

40. *Bemis*, 427 N.E.2d at 1062.

41. *Burton v. L.O. Smith Foundry Prods. Co.*, 529 F.2d 108, 111 (7th Cir. 1976).

42. K. ROSS & B. WRUBEL, PRODUCT LIABILITY 1989: WARNINGS, INSTRUCTIONS, AND RECALLS 47-48 (1989).

43. R. EPSTEIN, *supra* note 39, at 145.

44. *Nichols v. Union Underwear Co.*, 602 S.W.2d 429, 432 (Ky. 1980).

45. *Id.*

46. W. KEETON, PROSSER & KEETON ON TORTS § 99 (5th ed. 1984).

47. *Id.*

the doctrine by some courts outside of the state, Indiana continued to follow the open and obvious danger rule until the *Koske* decision.

IV. THE *KOSKE V. TOWNSEND ENGINEERING* DECISION

In *Koske*, the court made several statements that are both troubling and potentially confusing for the future status of product liability in Indiana.

A. The Court's Interpretation of "Defective Condition" and "Unreasonably Dangerous"

Section 402A and the Indiana Product Liability Act define "unreasonably dangerous" as dangerous "to an extent beyond that which would be *contemplated by the ordinary consumer who purchases it.*"⁴⁸ However, the *Koske* court shifted the focus of inquiry from the consumer to the product and the manufacturer. It stated that "[t]he concepts of 'defective condition' and 'unreasonably dangerous' focus the relevant inquiry upon the product and the manufacturer or seller, as assessed by an objective standard, regarding expected use."⁴⁹ This changes the focus from what the ordinary consumer expects to what the ordinary manufacturer expects.

Next, the *Koske* court declared that the language of the open and obvious danger rule, as formulated in *Bemis Co. v. Rubush*⁵⁰ and applied in other cases, exceeded the meaning of "defective condition" and "unreasonably dangerous."⁵¹ "By precluding liability whenever the defect is open and obvious, patent, or not hidden, [the rule] tended to obscure or minimize consideration of human factors related to the foreseeable circumstances of expected product use."⁵²

To illustrate how the rule obscures or minimizes consideration of human factors, the supreme court pointed to the results of three Indiana Court of Appeals decisions. The court of appeals in *Koske* held that summary judgment in favor of the defendant was proper because there were no genuine issues of fact.⁵³ In *FMC Corp. v.*

48. IND. CODE § 33-1-1.5-2 (1988) (emphasis added).

49. *Koske v. Townsend Eng'g*, 551 N.E.2d 437, 440 (Ind. 1990).

50. For the language of the rule as formulated in *Bemis*, see *supra* text accompanying note 29.

51. *Koske*, 551 N.E.2d at 441.

52. *Id.*

53. *Koske v. Townsend Eng'g*, 526 N.E.2d 985, 992 (Ind. Ct. App. 1980), *rev'd*, 551 N.E.2d 437 (Ind. 1990). The appellate court acknowledged that the relevant danger is not necessarily the injury-producing mechanism of a machine. Margaret argued that the relevant danger was the slipping of the jowls when used to push other jowls into the blades. Even if this was the relevant danger, the appellate court concluded that it was open and obvious because Margaret knew that the conveyor was wet and that the jowls were stiff, cold, and icy. Common experience would alert her to the risk of the jowl slipping as it approached the blades.

Brown,⁵⁴ the court of appeals held that providing an instruction on the open and obvious danger rule was reversible error because the trial court instructed the jury that the rule was an affirmative defense, thus shifting the burden of proof to the defendant.⁵⁵ In *Miller v. Todd*,⁵⁶ the plaintiff alleged that the open and obvious danger rule should not relieve a motorcycle manufacturer from the duty to design a crashworthy vehicle. The court of appeals stated that as a matter of law, the absence of a crash bar on a motorcycle is an open and obvious danger to the ordinary consumer.⁵⁷ Although the court cited the appellate decisions of *Koske*, *FMC Corp.*, and *Miller* as cases minimizing consideration of human factors, it referred to *Hoffman v. E.W. Bliss Co.*,⁵⁸ *Kroger Co. Sav-On Store v. Presnell*,⁵⁹ and *Corbin v. Coleco Industries*,⁶⁰ as cases in which human factors were not

54. 526 N.E.2d 719 (Ind. Ct. App. 1988), *rev'd*, 551 N.E.2d 444 (Ind. 1990).

55. *Id.* at 728. The court agreed with the plaintiff that denial of summary judgment was proper. The court explained that "whether a danger is open and obvious depends not just on what people can see with their eyes but also on what they know and believe about what they see." *Id.* at 725 (quoting *Corbin v. Coleco Indus.*, 748 F.2d 411 (7th Cir. 1984)). The court found that reasonable men believe it is safe to operate a crane near power lines if the crane is 10 to 15 feet away from the line.

56. 518 N.E.2d 1124 (Ind. Ct. App. 1988), *rev'd*, 551 N.E.2d 1139 (Ind. 1990).

57. *Id.* at 1126. The court did not reach the issue of whether Indiana adheres to the doctrine of crashworthiness. Instead, the focus was on whether the manufacturer provided a product in a defective condition unreasonably dangerous to the user. Applying an objective standard, the court concluded that an ordinary consumer is aware that the absence of a crash bar affords no protection to the legs of an unenclosed rider.

58. 448 N.E.2d 277 (Ind. 1983). It is one thing to excuse a manufacturer from liability for injuries caused by dangers which are open and obvious. However, excusing liability is not appropriate when the injury is caused by mechanisms that, due to a hidden defect, cause it to operate or malfunction at a time when the user has every reason to expect that it will not. The court could not say as a matter of law that the plaintiff's injury was caused by a patent defect. Evidence existed that tended to show that the descent of the ram in the metal punch press may have been caused by either a true double trip or an uninitiated spontaneous cycle of the press. *Id.* at 285.

59. 515 N.E.2d 538 (Ind. Ct. App. 1987). The plaintiff alleged that a lounge chair was defective and unreasonably dangerous because the defendants failed to provide instructions or warnings about how to open the chair. She argued that without proper instructions or warnings, the danger of the chair's collapse was not apparent. The court affirmed the trial court's denial of summary judgment. It stated that the danger posed to the consumer was not open and obvious as a matter of law. The court held that it was for the jury to decide whether the danger was patent. *Id.* at 543.

60. 748 F.2d 411 (7th Cir. 1984). Plaintiff injured himself when he hit his head on the bottom after diving into an above-ground swimming pool manufactured by the defendant. Expert testimony indicated that people are generally aware of the danger of diving into shallow water, but they believe there is a safe way to do it (*i.e.* by executing a flat shallow dive). Thus, the risk of spinal cord injury from diving into shallow water cannot be an open and obvious danger as a matter of law. The court concluded that "whether a danger is open and obvious depends not just on what people can see with their eyes but also on what they know and believe about what they see." *Id.* at 417-18.

minimized or obscured.⁶¹

B. *Obviousness as a Factor of Incurred Risk*

Next, the *Koske* court stated that the proper use of the obviousness of a danger is as “an appropriate consideration in product strict liability to evaluate the actual state of mind of the product user when the affirmative defense of incurred risk is asserted.”⁶² If the plaintiff establishes that the product was sold in a defective condition unreasonably dangerous to the user, the defendant may still avoid liability by showing that the plaintiff had *actual* knowledge and appreciation of the specific danger and voluntarily accepted the risk.⁶³ The use of “obviousness” in this manner involves an inquiry into the consumer’s subjective state of mind. The question becomes “what was this plaintiff’s awareness” instead of “what is the ordinary consumer’s awareness.”

The *Koske* court criticized cases that invoke an objective test when considering the obviousness of a danger.⁶⁴ The court stated that:

Many subsequent cases applied the *Bemis* open and obvious danger language not merely to aid in the determination of “unreasonably dangerous” relative to a product and its manufacturer, but also to engraft upon § 402A an additional element involving evaluation of the plaintiff’s conduct separate and apart from the affirmative defense of incurred risk.⁶⁵

61. *Koske v. Townsend Eng’g*, 551 N.E.2d 437, 441 (Ind. 1990).

62. *Id.* See also IND. CODE § 33-1-1.5-4(1) (1988) (“it is a defense that the user or consumer bringing the action knew of the defect and was aware of the danger and nevertheless proceeded unreasonably to make use of the product and was injured by it”); RESTATEMENT (SECOND) OF TORTS § 496A (1965) (“a plaintiff who voluntarily assumes a risk of harm arising from the negligent or reckless conduct of the defendant cannot recover for such harm”); RESTATEMENT (SECOND) OF TORTS § 402A comment n (1965) (voluntarily proceeding to encounter a known danger is a defense under § 402A).

63. *Koske*, 551 N.E.2d at 441.

64. *Id.*

65. *Id.* The court criticized *Angola State Bank v. Butler Mfg.*, 475 N.E.2d 717 (Ind. Ct. App. 1985) (“notwithstanding evidence showing unguarded chain and sprocket mechanism was unreasonably dangerous, the open and obvious danger rule did apply to preclude manufacturer liability”); *Ragsdale v. K-Mart Corp.*, 468 N.E.2d 524 (Ind. Ct. App. 1984) (“summary judgment upheld applying open and obvious danger rule as proper consideration for determining whether plaintiff acted reasonably in exposing himself to danger”); *Law v. Yukon Delta, Inc.*, 458 N.E.2d 677 (Ind. Ct. App. 1984) (“open and obvious danger rule applies objective test to determine whether plaintiff should have recognized the danger”); and *Bryant-Poff, Inc. v. Hahn*, 454 N.E.2d 1223 (Ind. Ct. App. 1982), *cert. denied*, 465 U.S. 1075 (1984) (“rule requires objective test of what the user should have known to determine if an unshielded power takeoff shaft was open and obvious”).

Furthermore, the use of an objective standard was held to be inappropriate because it is similar to the defense of contributory negligence which is not an available defense in strict liability in tort.⁶⁶

C. Statutory Language

To further support its statement that the open and obvious danger rule no longer applies in strict liability cases, the court analyzed the purpose and language of the 1978 Indiana Product Liability Act. The Act only undertook a “Codification and Restatement” of strict liability.⁶⁷ The court noted that the Act enumerates the affirmative defenses applicable to strict liability in tort, but does not attempt to codify and restate defenses applicable to claims based on negligence.⁶⁸ “Because of the express intention to codify and restate, and because the resulting enactment comprehensively addressed the subject matter, [the court] conclude[d] that with the 1978 Product Liability Act the legislature entered, occupied, and preempted the field of product strict liability in tort.”⁶⁹

Determining whether the *Bemis* open and obvious danger rule is included in the statute depends on whether the Act, by express terms or by unmistakable implication, made changes in the preexisting common law.⁷⁰ The court found an unmistakable implication that the *Bemis* rule was excluded from the Act’s codification and restatement of strict liability law:

The Act not only employed the language of Restatement (Second) § 402A without explicitly incorporating the words open and obvious or requiring that a defect be latent or concealed, but it also expressly delineated the allowable defenses to strict liability in tort to include evaluation of the product user’s conduct only by a subjective rather than an objective standard.⁷¹

Therefore, the court held that the Indiana open and obvious danger rule does not apply to strict liability claims under the Indiana Product Liability Act.⁷²

66. *Koske v. Townsend Eng’g*, 551 N.E.2d 437, 441 (Ind. 1990). See IND. CODE § 33-1-15-4 (1988) (enumerating defenses to strict liability in tort).

67. *Koske*, 551 N.E.2d at 442.

68. *Id.*

69. *Id.*

70. *Id.*

71. *Id.*

72. *Id.*

D. The Open and Obvious Danger Rule: Applicability to Negligence Theory of Product Liability

Indiana common law recognizes that the open and obvious danger rule is applicable to product liability claims grounded in negligence.⁷³ The Indiana Supreme Court offered three reasons why the open and obvious danger rule is applicable to product liability claims based on negligence. First, the Product Liability Act's express codification and restatement of the common law does not extend to general product negligence law.⁷⁴ Second, the Act does not comprehensively cover the subject matter of general product negligence law.⁷⁵ Third, the Act does not by express terms or unmistakable implication change general product negligence law.⁷⁶ Therefore, the common-law open and obvious danger rule of *Bemis* and its progeny is not superseded by the Act in product negligence liability cases.⁷⁷

V. ANALYSIS OF *KOSKE V. TOWNSEND ENGINEERING*

A. “*Defective Condition*” and “*Unreasonably Dangerous*”

The *Koske* court's misunderstanding of the open and obvious danger rule is illustrated by its treatment of the terms “defective condition” and “unreasonably dangerous.” The court stated that the two concepts “focus the relevant inquiry on the product and the manufacturer or seller, as assessed by an objective standard, regarding expected use.”⁷⁸ As defined by section 402A and by the Indiana statute, however, the relevant inquiry is on the ordinary consumer and his expectations.⁷⁹ By shifting the focus of the inquiry from the *ordinary consumer* to the ordinary manufacturer,⁸⁰ the court, without expressly stating so, rejected the consumer expectation test.

The consumer expectation test used by Indiana courts at common law was incorporated into the Product Liability Act. The statute refers to unreasonably dangerous as “any situation in which the use of a product exposes the user or consumer to a risk of physical harm to an extent beyond that *contemplated by the ordinary consumer* who purchases

73. *Id.* at 443. See *Bridgewater v. Economy Eng'g Co.*, 486 N.E.2d 484 (Ind. 1985).

74. *Koske v. Townsend Eng'g*, 551 N.E.2d 437, 443 (Ind. 1990).

75. *Id.*

76. *Id.*

77. *Id.*

78. *Id.* at 440.

79. See *supra* notes 11, 21 and accompanying text.

80. See *supra* note 49 and accompanying text.

it with the ordinary knowledge about the product's characteristics common to the community of consumers."⁸¹ This language is similar to the language of comment i.⁸² Failing to properly apply the consumer expectation test included within the product liability statute was the first step leading to the court's misunderstanding of the open and obvious danger rule.

As applied in a number of jurisdictions, the consumer expectation test incorporates the open and obvious danger rule by precluding a determination of product defect when the danger associated with the product is patent and thus, not more dangerous than the ordinary consumer would expect.⁸³ In explaining this relationship, the Fifth Circuit Court of Appeals stated:

The key to the *Manitowoc*⁸⁴ decision was the court's determination that consumer expectations should be the primary focus in determining whether a product is unreasonably dangerous. . . . Under this analysis, a product design incorporating an open and obvious hazard could never be unreasonably dangerous, because it could never be more dangerous than an ordinary consumer would expect.⁸⁵

Both the consumer expectation test and the open and obvious danger rule approach a product defect from the standpoint of the *ordinary consumer*. Even though both doctrines are criticized by courts and commentators, by incorporating section 402A into the Product Liability Act, the Indiana legislature intended for the consumer expectation test to be the standard used for determining defectiveness in strict product liability cases. Because of the interrelatedness of the two concepts, the court's decision in *Koske* is illogical.

Next, the *Koske* court asserted that the language of the open and obvious danger rule exceeds the meaning of "defective condition" and

81. IND. CODE § 33-1-1.5-2 (1988) (emphasis added).

82. "The article sold must be dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it, with the ordinary knowledge common to the community as to its characteristics." RESTATEMENT (SECOND) OF TORTS, § 402A comment i (1965).

83. K. ROSS & B. WRUBEL, PRODUCT LIABILITY OF MANUFACTURERS 1988: PREVENTION AND DEFENSE 11-12 (1988).

84. *Gray v. Manitowoc Co.*, 771 F.2d 866 (5th Cir. 1985). The plaintiff was struck by the boom of a construction crane manufactured by the defendant. The plaintiff alleged that the operator's vision was obstructed by the boom and that mirrors, closed-circuit television, or other devices should have been provided to enable the operator to see blind spots. The defendant argued that the lack of devices did not render the crane defective and that the hazards of operation were open and obvious to ordinary users of the crane.

85. *Melton v. Deere & Co.*, 887 F.2d 1241, 1248 (5th Cir. 1989) (citation omitted).

"unreasonably dangerous" because it "obscures or minimizes consideration of human factors related to the foreseeable circumstances of expected product use."⁸⁶ Although the court did not expound upon this statement, the inference is that the court was displeased that the rule operates as a total bar to recovery by allowing summary judgment in favor of the defendant or by placing the burden of proving latency upon the plaintiff. This conclusion is drawn from the cases cited by the court.⁸⁷ However, these cases discredit the court's assertion, rather than support it.

The court seems to divide these cases into two irreconcilable groups: (1) those in which human factors are minimized or obscured⁸⁸ and (2) those in which human factors are given consideration.⁸⁹ Rather than viewing these two groups of cases as standing for dissimilar propositions, one can view them as harmonious. *Hoffman*, *Kroger*, and *Corbin* illustrate the courts' efforts since *Bemis* to refine the proper application of the open and obvious danger rule.

In *Hoffman*, the court explained that while the injury-producing mechanism of the machine (the descent of a ram on a metal punch press) was open and obvious, the true defect or danger was the internal malfunction that caused the ram to descend without warning or without activation by the operator.⁹⁰ Thus, the *Hoffman* court provided clearer guidelines for identifying the relevant danger. Likewise, in *Corbin*, the court narrowed the meaning of open and obvious by stating that "[w]hether a danger is open and obvious depends not just on what people can see with their eyes but also on what they know and believe about what they see."⁹¹ Finally, in *Kroger*, the court determined that the danger was not always open and obvious *as a matter of law*.⁹²

The appellate decisions of *Koske*, *Miller*, and *FMC Corp.* did not contradict the propositions announced in *Hoffman*, *Corbin*, or *Kroger*.⁹³ Rather, *Koske* and *Miller* exemplified situations in which patency was appropriately a matter of law. In fact, the *Koske* court acknowledged the proposition from *Hoffman* that the relevant danger is not necessarily

86. *Koske v. Townsend Eng'g*, 551 N.E.2d 437, 441 (Ind. 1990).

87. See *supra* notes 53-60 and accompanying text.

88. See *supra* note 52-57 and accompanying text.

89. See *supra* notes 62-63 and accompanying text.

90. *Hoffman v. E.W. Bliss Co.*, 448 N.E.2d 277, 285 (Ind. 1983).

91. *Corbin v. Coleco Indus.*, 748 F.2d 411, 417-18 (7th Cir. 1984).

92. *Kroger Co. Sav-On Store v. Presnell*, 515 N.E.2d 538, 543 (Ind. Ct. App. 1987).

93. It should be noted that the appellate decisions in *FMC Corp.* and *Miller* were reversed by the Indiana Supreme Court *after* it rendered its decision in *Koske*. Hence, the Indiana Supreme Court was applying the new proposition of *Koske* that the open and obvious danger rule is inapplicable to strict liability claims.

the injury-producing mechanism of a machine.⁹⁴ However, the court distinguished the facts in *Koske* from those in *Hoffman*. The court of appeals concluded that even if “the relevant danger is the slipping of one jowl when used to push other jowls into the blades, the facts are not in conflict that the danger was open and obvious.”⁹⁵

In addition, *FMC Corp.* illustrates that refinements made by prior courts were applied in subsequent cases. The appellate court in *FMC Corp.* affirmed the trial court’s denial of the defendant’s motion for summary judgment.⁹⁶ The court cited *Corbin* for the proposition that “whether a danger is open and obvious depends not just on what people can see with their eyes but also on what they know and believe about what they see.”⁹⁷ The *Koske* court probably viewed *FMC Corp.* negatively because the *FMC Corp.* court held that providing an instruction that classifies the open and obvious danger rule as an affirmative defense is reversible error.⁹⁸ The *Koske* court’s disapproval is understandable because the court also mistakenly classified “obviousness” as a factor of the affirmative defense of incurred risk.⁹⁹

B. *Obviousness as an Element in Evaluating the User’s Actual State of Mind*

The *Koske* court held that “obviousness” could properly be used as an aid in evaluating the user’s actual state of mind in relation to the affirmative defense of incurred risk.¹⁰⁰ The court rejected the objective test because it is akin to contributory negligence, which is not an available defense under the statute.¹⁰¹ The court’s analysis rests on the erroneous assumption that the patent danger rule is a defense. This error is central to the court’s holding.

The *Bemis* court held that “to impress liability upon manufacturers, the defect must be hidden and not normally observable.”¹⁰² Thus, the question is not whether liability will be excused because of an affirmative defense, but rather whether it will be imposed in the first place. One element of the plaintiff’s *prima facie* case is to show that the defect is

94. *Koske v. Townsend Eng’g*, 526 N.E.2d 985, 992 (Ind. Ct. App. 1988), *rev’d*, 551 N.E.2d 437 (Ind. 1990).

95. *Id.*

96. *FMC Corp. v. Brown*, 526 N.E.2d 719, 728 (Ind. Ct. App. 1988).

97. *Id.* at 725.

98. *Id.*

99. See *supra* note 62 and accompanying text.

100. *Id.*

101. See *supra* note 66 and accompanying text.

102. *Bemis Co. v. Rubush*, 427 N.E.2d 1058, 1061 (Ind. 1981).

latent. This is distinctly different from the affirmative defense of contributory negligence.

Using an objective standard does not engraft an additional element on section 402A.¹⁰³ First, section 402A specifically calls for an objective standard. Comment i states that for the rule to apply, the product "must be dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it."¹⁰⁴ The key phrase is "beyond the contemplation of the ordinary consumer." In other words, comment i implies that a danger must be latent or hidden. Even though the language of section 402A and the comments do not expressly use the terms "open and obvious" or require that a defect be hidden, a reasonable inference is that a latent defect is required before strict liability will be imposed. This inference is supported by both the language of the comment and by the interpretation made by courts using section 402A.¹⁰⁵

C. *The 1978 Indiana Product Liability Act*

The *Koske* court stated that the 1978 Indiana Product Liability Act is a codification and restatement of the law.¹⁰⁶ To support its conclusion that the patent danger rule is inapplicable to cases based on strict liability, the court stated that the Act enumerates the available affirmative defenses without including the open and obvious doctrine.¹⁰⁷ This lends little support to the court's conclusion because the doctrine is not an affirmative defense. Instead, it is used to determine whether a plaintiff has established his *prima facie* case by showing that the danger or defect was latent or hidden.

The second argument that the court offered is that the Act, by unmistakable implication, changed the preexisting common law because the Act employs the language of section 402A without explicitly incorporating the words "open and obvious" or requiring that a defect be latent.¹⁰⁸ The "unmistakable implication" that the court draws is not so clear. Neither section 402A nor its comments use the words "open and obvious" or "latent."¹⁰⁹ These terms are implied by the language "must be dangerous to an extent beyond that which would be contemplated by the ordinary consumer" in comment i.¹¹⁰

103. See *supra* note 65 and accompanying text.

104. RESTATEMENT (SECOND) OF TORTS § 402A comment i (1965).

105. See *Bemis*, 427 N.E.2d at 1058.

106. *Koske v. Townsend Eng'g*, 551 N.E.2d 437, 442 (Ind. 1990).

107. *Id.*

108. *Id.*

109. See RESTATEMENT (SECOND) OF TORTS § 402A (1965).

110. *Id.* § 402A comment i.

*D. Open and Obvious Danger Rule and Product Liability Claims
Based on Negligence*

The final holding of *Koske* is that the common-law open and obvious danger rule of *Bemis* is applicable in product liability claims based on negligence. This presents a problem for future application of the rule. Because "a pleading may set forth two [2] or more statements of a claim . . . alternatively or hypothetically,"¹¹¹ plaintiffs often advance both strict liability and negligence theories in product liability cases. However, plaintiffs may have difficulty distinguishing between the two theories. In fact, strict liability under section 402A rarely leads to a different conclusion than that drawn under the laws of negligence.¹¹² The *Koske* decision requires plaintiffs to make a distinction between two theories which may be easy in theory, but difficult in practice.

In addition, there is great potential for confusion when instructing the jury on each theory under the rule in *Koske*. A defendant will be entitled to an instruction of the open and obvious danger rule regarding the negligence theory, but not the strict liability theory. To expect the jury to apply the open and obvious danger rule with respect to negligence, but to refrain from using it in regard to strict liability, is wishful thinking. The problem is further compounded by the fact that Indiana courts do not allow special verdict forms.¹¹³ This frustrates the ability to determine the basis of the jury's verdict.

The *Koske* court stated that the *Bemis* rule is applicable to product liability claims based on negligence because the Act does not extend to negligence claims and does not change the general product negligence law.¹¹⁴ Although the *Koske* court misinterpreted the *Bemis* rule, it nonetheless recognized it as applying in product negligence cases. As the preceding analysis suggests, no rationale exists for distinguishing between the use of the open and obvious danger rule in negligence and strict liability theories of product liability.

VI. CONCLUSION

Although the *Koske* court attempted to address the problems of the open and obvious danger rule, it misinterpreted case law and statutory law surrounding strict products liability. The rule's critics may have valid concerns regarding the impact of the rule. The main thrust of this Note, however, is that the consumer expectation test is so interrelated

111. IND. TR. R. 8(E)(2).

112. R. CARTWRIGHT & J. PHILLIPS, *supra* note 14, at 511.

113. IND. TR. R. 49 (special verdicts and interrogatories to the jury are abolished).

114. *Koske v. Townsend Eng'g*, 551 N.E.2d 437, 443 (Ind. 1990).

with the open and obvious danger rule that one doctrine cannot be discarded without affecting the other. As long as the statute calls for the consumer expectation standard to be used in deciding whether a product is in a defective condition unreasonably dangerous, then the open and obvious danger rule should also be followed.

Certain steps can be taken to prevent the injustices feared by the court and critics of the rule. First, courts can embrace the idea that the openness and obviousness of a danger may be a question of fact for the trier to decide. The *Bemis* court did not hold that patenty was a pure question of law.¹¹⁵ Later cases made this clear.¹¹⁶ The Indiana Supreme Court stated that “[i]n *Bemis*, we did not hold that the question of whether an alleged danger is open and obvious is a matter of law in *all* cases.”¹¹⁷ However, if from the uncontested facts no reasonable jury properly instructed in Indiana law could infer that the danger was not patent, then summary judgment is proper.¹¹⁸

Critics argue that the open and obvious danger rule encourages manufacturers to leave off safety devices in order to make dangers patent.¹¹⁹ As one justice observed, holding a danger open and obvious *as a matter of law* may be the factor causing manufacturers to leave off safety devices.¹²⁰ This problem may be reduced if the question of patenty is left to the trier of fact. Defendants may feel less comfortable leaving this determination to the jury, thus weakening the incentive to leave off safety devices.

A second step that courts could take to preserve the value of the doctrine is to establish clearly defined exceptions. One example is when a machine “invites” the user into its zone of danger.¹²¹ In *Berg v. Sukup Co.*,¹²² a grain farmer injured his left arm when it became entangled in the rotating shaft of a grain drying system. The system consisted of a

115. The court stated that a machine may not be built with flimsy parts concealed by an exterior such as to mislead a user into believing it is safe and stable when, in fact, it is not. Therefore, whether there is a concealed defect or hidden danger to a user is a question of fact. *Bemis Co. v. Rubush*, 427 N.E.2d 1058, 1062 (Ind. 1981).

116. See *Bridgewater v. Economy Eng'g*, 486 N.E.2d 484 (Ind. 1985); *Hoffman v. E.W. Bliss Co.*, 448 N.E.2d 277 (Ind. 1983); *Kroger Co. Sav-On Store v. Presnell*, 515 N.E.2d 538 (Ind. Ct. App. 1987).

117. *Bridgewater*, 486 N.E.2d at 488 (emphasis in original).

118. *Estrada v. Schmutz Mfg. Co.*, 734 F.2d 1218, 1220 (8th Cir. 1984).

119. See J. BEASLEY, *supra* note 15, at 91; R. CARTWRIGHT & J. PHILLIPS, *supra* note 14, § 5.33.

120. *Bryant-Poff, Inc. v. Hahn*, 453 N.E.2d 1171, 1174 (Ind. 1983) (Hunter, J., dissenting).

121. K. ROSS & B. WRUBEL, *supra* note 42, at 52 (citing *Berg v. Sukup Mfg. Co.*, 355 N.W.2d 833 (S.D. 1984)).

122. 355 N.W.2d 833 (S.D. 1984).

grain bin with several horizontal and vertical augers designed to circulate the grain. The horizontal auger was fitted with a slide gate which was customarily used to secure samples of processed grain. An unshielded drive shaft was situated four to five inches from the slide gate. Although other means of obtaining grain samples existed, the slide gate was the most feasible method. While obtaining a grain sample from the slide gate, Berg entangled his sleeve in the rotating shaft and injured his arm.

The court found that Berg was using the system in a manner foreseen and expected by Sukup.¹²³ "The slide gate location actually invited the operator into the location of the danger."¹²⁴ In answer to the defendant's argument that the danger was open and obvious, the court declared that the reasonableness of the plaintiff's conduct was a factor for the jury to consider.¹²⁵

Another measure would be to distinguish between design cases and warning cases. The open and obvious danger rule is more applicable to warning cases. "It is well established that there is no duty resting upon a manufacturer or seller to warn of a product-connected danger which is obvious or generally known. . . . *The same rule applies when it appears that the person using the product should know of the danger.*"¹²⁶ The purpose of a warning is to apprise a party of a danger of which he is unaware, thereby enabling him to protect himself against the danger.¹²⁷ When the danger is fully obvious and appreciated, no value is added by issuing a warning.¹²⁸ If suppliers are required "to warn of all obvious dangers inherent in a product, '[t]he list of foolish practices warned against would be so long it would fill a volume.'"¹²⁹

The most drastic measure that could be taken to avoid the disadvantages of the open and obvious danger rule is for the legislature to adopt a new test for determining when a product is defective. Many jurisdictions have adopted a risk-utility approach which balances the risks associated with the product and the utility of the product.¹³⁰ Other jurisdictions combine the consumer expectation test with the risk-utility

123. *Id.* at 836.

124. *Id.*

125. *Id.* at 835-36.

126. American Optical Co. v. Weidenhamer, 457 N.E.2d 181, 188 (Ind. 1983) (emphasis in original).

127. K. Ross & B. WRUBEL, *supra* note 42, at 47.

128. *Id.* at 47-48.

129. Plante v. Hobart Corp., 771 F.2d 617, 620 (1st Cir. 1985) (quoting Kerr v. Koemm, 557 F. Supp. 283, 288 n.2 (S.D.N.Y. 1983)).

130. See, e.g., Ortho Pharmaceutical Corp. v. Heath, 722 P.2d 410 (Colo. 1986); Bilotta v. Kelley Co., 346 N.W.2d 616 (Minn. 1984); Suter v. San Angelo Foundry & Mach., 81 N.J. 150, 406 A.2d 140 (1979); General Motors Corp. v. Turner, 584 S.W.2d 844 (Tex. 1979).

test.¹³¹ However, one court has held that the risk-utility test is proper only when the consumer expectation test is inappropriate.¹³²

Even though other approaches are available, the fact remains that the Indiana Product Liability Act cloaks "defective condition" and "unreasonably dangerous" in the garb of the consumer expectation test.¹³³ In doing so, the Indiana legislature impliedly affirmed the use of the common-law open and obvious danger rule. Because these two concepts are inherently connected, the court in *Koske* erred when it held that the open and obvious danger rule is inapplicable to product liability claims based on strict liability.

JODIE L. MINER

131. See, e.g., *Dart v. Wiebe Mfg.*, 147 Ariz. 242, 709 P.2d 876 (1985); *Knitz v. Minster Mach.*, 69 Ohio St. 2d 460, 432 N.E.2d 814 (1982).

132. *Sours v. General Motors Corp.*, 717 F.2d 1511 (6th. Cir. 1983).

133. IND. CODE §§ 33-1-1.5-2 to -2.5 (1988).

“Lawsuit For Sale”: An Analysis of the Assignability of Legal Malpractice Claims

An issue with potentially far-reaching implications for all practicing attorneys was addressed in a recent Indiana Court of Appeals decision. In *Picadilly, Inc. v. Raikos*¹ the plaintiff-assignee, Charles Colvin, was injured in an automobile accident with a drunk driver who became intoxicated while patronizing Picadilly’s bar. Colvin brought a dram shop action against Picadilly and was awarded \$150,000 in punitive damages. While the appeal of the judgment of the dram shop action was pending, Picadilly filed a legal malpractice complaint against its attorneys, Raikos and Thomas. The theory in the malpractice action was “failure to exercise proper care in defending Picadilly [in the dram shop trial], ‘including but not limited to a failure to properly preserve any objection to the Court’s improper jury instructions on punitive damages.’”²

Shortly after the court of appeals affirmed the punitive damage award in the dram shop action, the trial court entered summary judgment in favor of the attorneys in the legal malpractice action and the attorneys appealed. Subsequently, a United States Bankruptcy Court entered an order confirming Picadilly’s reorganization plan under Chapter 11 bankruptcy. The plan included a full discharge of the punitive damages debt to Colvin on the condition that Picadilly pay Colvin \$5,000 and transfer its legal malpractice claim against Raikos and Thomas to him. The trial court’s decision in the malpractice claim was on appeal at the time of the assignment. Consequently, when the legal malpractice action appeared before the court of appeals, Charles Colvin, as assignee, was the plaintiff.

Although the court of appeals affirmed the summary judgment in favor of the attorneys,³ this Note focuses on the assignability of the malpractice claim. The issue of whether legal malpractice claims are assignable was one of first impression in Indiana.⁴ After reviewing the arguments that have been advanced in other jurisdictions, the court of appeals held that such an assignment is valid under the circumstances.⁵

The purpose of this Note is to evaluate the various arguments and considerations regarding the assignability of legal malpractice claims.

1. 555 N.E.2d 167 (Ind. Ct. App. 1990), *argued*, No. 41A01-8908-CV-311 (Ind. Feb. 20, 1991).

2. *Id.*

3. *Id.* at 170.

4. *Id.* at 168.

5. *Id.* at 169.

Section I reviews the background of legal malpractice and discusses the relevant characteristics of malpractice actions. Section II discusses the law of assignment and identifies the factors employed in determining whether a particular action is assignable. Arguments advanced by courts opposing the assignability of legal malpractice actions are described in Section III, and the arguments in support of these assignments are identified in Section IV. Section V proposes a resolution to the controversy surrounding this issue. This Section also suggests that encouraging the assignment of the judgment, rather than the legal malpractice action, is a viable compromise worthy of serious consideration.

I. THE LAW OF LEGAL MALPRACTICE

In order to recover on a claim of legal malpractice, a litigant must plead and prove every essential element thereof: duty, breach of duty, proximate cause, and damages.⁶ With respect to the first element, a contractual relationship between the attorney and client is the usual basis for establishing a duty owed by the attorney.⁷ It is possible, however, for nonclients to have standing to sue for legal malpractice. Many jurisdictions permit a plaintiff to assert a claim against an attorney despite the lack of an attorney-client relationship. In these jurisdictions, the duty owed by an attorney does not depend upon privity of contract between the plaintiff and the attorney.⁸ Furthermore, several courts allow nonclients to assert legal malpractice claims by permitting the assignment of such actions.⁹ These theories of extended liability stem from the erosion of the traditional theory that privity of contract is required to show that a duty is owed by the attorney.

A. Development of the Privity Requirement

Historically, courts did not allow the assertion of negligence actions in the absence of privity. The origin of this common-law rule dates back to the 1842 English decision, *Winterbottom v. Wright*.¹⁰ In *Winterbottom*, a coach manufacturer was found not liable to a driver who suffered personal injuries because of a manufacturing defect.¹¹ The court

6. D. HORAN & G. SPELMIRE, ATTORNEY MALPRACTICE: PREVENTION AND DEFENSES 11-1 (1987). See also *Goodley v. Wank & Wank, Inc.*, 62 Cal. App. 3d 392, 394, 133 Cal. Rptr. 83, 85 (1976); *Fiddler v. Hobbs*, 475 N.E.2d 1172, 1173 (Ind. Ct. App. 1985).

7. See, e.g., *McGlone v. Lacey*, 288 F. Supp. 622 (D.S.D. 1968); *Bloomer Amusement Co. v. Eskanazi*, 75 Ill. App. 3d 117, 394 N.E.2d 16 (1979).

8. See *infra* notes 33-43 and accompanying text.

9. See *infra* notes 119-23.

10. 152 Eng. Rep. 402 (1842).

11. *Id.* at 405.

reached its conclusion because privity of contract did not exist between the manufacturer and the driver:

If we were to hold that the plaintiff could sue in such a case, there is no point at which such actions would stop. The only safe rule is to confine the right to recover to those who enter into the contract: if we go one step beyond that, there is no reason why we should not go fifty. The only real argument in favour of the action is that this is a case of hardship; but that might have been obviated, if plaintiff had made himself a party to the contract.¹²

The privity requirement in the context of an attorney malpractice action was first addressed in *National Savings Bank v. Ward*.¹³ The attorney in *Ward* negligently overlooked a previously recorded deed when he examined the title to real estate that was offered by his client as collateral for a loan. After the client defaulted on the loan payments, the bank tried unsuccessfully to obtain the property. The bank was subsequently precluded from recovering damages from the attorney even though it relied on his performance. The United States Supreme Court held that a legal malpractice action cannot be maintained by someone outside the attorney-client relationship, even if the third party's injury was proximately caused by the attorney's negligence.¹⁴ As the Supreme Court stated:

Beyond all doubt, the general rule is that the obligation of the attorney is to his client and not to a third party, and unless there is something in the circumstances of this case to take it out of that general rule, it seems clear that the proposition [that there can be no liability in the absence of privity] of the defendant [attorney] must be sustained.¹⁵

The Supreme Court also acknowledged that neither fraud nor collusion was alleged by the plaintiff.¹⁶ This recognition suggested an exception to the strict privity rule.

Currently, the rule followed in many states is that, in the absence of fraud or collusion, privity of contract is required to maintain a cause of action for legal malpractice.¹⁷ Other states have expanded the scope

12. *Id.*

13. 100 U.S. 195 (1879).

14. *Id.*

15. *Id.* at 200.

16. *Id.* at 199.

17. See, e.g., *Pelham v. Greisheimer*, 92 Ill. 2d 13, 19, 440 N.E.2d 96, 99 (1982)

of the legal malpractice action, however, and have found liability in the absence of privity.¹⁸

B. Demise of the Privity Requirement

The privity requirement was first eroded in products liability actions. The court in *MacPherson v. Buick Motor Co.*¹⁹ was the first to expressly disregard the privity rule. In *MacPherson*, the plaintiff purchased a car from a retail dealer and was injured when one of the wheels fell off. The plaintiff sued the car manufacturer for negligent inspection. Although a contractual relationship did not exist between the parties, the manufacturer was found to owe a duty to the plaintiff:²⁰

We have put aside the notion that the duty to safeguard life and limb, when the consequences of negligence may be foreseen, grows out of contract and nothing else. We have put the source of the obligation where it ought to be. We have put its source in the law.²¹

The court reasoned that, by placing the product in the free flow of commerce, Buick Motor Company assumed responsibility based on the foreseeability of harm, rather than on contract principles.²²

Courts quickly followed suit in other areas of the law. For example, the New York Court of Appeals found liability in the absence of privity in the area of negligent misrepresentation in *Glanzer v. Shepard*.²³ The court in *Glanzer* held a public weigher liable to a buyer who relied on a negligently issued certificate of weight, even though the seller hired

(“[t]he concept of privity has long protected attorneys from malpractice claims by non-clients”); *Flaherty v. Weinberg*, 303 Md. 115, 121, 492 A.2d 618, 620 (1985) (a “majority of American courts evidently continue to adhere to the view expressed . . . that absent fraud, collusion, or privity of contract, an attorney is not liable to a third party for professional malpractice”); *Clagett v. Dacy*, 47 Md. App. 23, 420 A.2d 1285 (1980); *Eustis v. David Agency, Inc.*, 417 N.W.2d 295, 298 (Minn. Ct. App. 1987) (“an attorney will not be liable to a non-client third-party for negligence. Liability arises only if that attorney acted with fraud, malice, or has otherwise committed an intentional tort.”); *Council Commerce Corp. v. Schwartz, Sachs & Kamhi, P.C.*, 144 A.D.2d 494, 534 N.Y.S.2d 1 (1988); *Scholler v. Scholler*, 10 Ohio St. 3d 98, 462 N.E.2d 158 (1984); *Dickey v. Jansen*, 73 S.W.2d 581, 582-83 (Tex. Ct. App. 1987) (“Texas authorities have consistently held that third parties have no standing to sue attorneys on causes of action arising out of their representation of others.”).

18. See *infra* notes 27-43 and accompanying text.

19. 217 N.Y. 382, 111 N.E. 1050 (1916).

20. *Id.* at 394, 111 N.E. at 1055.

21. *Id.* at 390, 111 N.E. at 1053.

22. *Id.*

23. 233 N.Y. 236, 135 N.E. 275 (1922).

the weigher to issue the certificate.²⁴ The court imposed liability in the absence of privity because the defendant-weigher knew that the buyer would rely on his certification.²⁵

The privity requirement has also been overcome in attorney malpractice actions in several states.²⁶ California was the first to hold an attorney liable for negligence in the absence of privity. In *Biakanja v. Irving*,²⁷ the Supreme Court of California allowed recovery by a third party who was an intended beneficiary of a will that was never probated because of improper attestation.²⁸ This holding was followed three years later in *Lucas v. Hamm*,²⁹ in which the Supreme Court of California stated:

As in *Biakanja*, one of the main purposes which the transaction between defendant and the testator intended to accomplish was to provide for the transfer of property to plaintiffs; the damage to plaintiffs in the event of invalidity of the bequest was clearly foreseeable; it became certain, upon the death of the testator without change of the will, that plaintiffs would have received the intended benefits but for the asserted negligence of defendant; and if persons such as plaintiffs are not permitted to recover for the loss resulting from negligence of the draftsman, no one would be able to do so, and the policy of preventing future harm would be impaired.³⁰

As a result of *Biakanja* and *Lucas*, attorneys are on notice of their potential liability to nonclients. *Heyer v. Flraig*,³¹ another will-drafting case in California, further defined this extended liability. In *Heyer*, the court expressly recognized that a separate and distinct duty is owed to an intended beneficiary who is afforded all the remedies available in tort.³² When such a duty arises, the attorney is liable for any breach which proximately caused the injuries suffered. Other jurisdictions followed California's lead and have found attorneys liable to nonclients when the action is based on one of the theories discussed below.

24. *Id.* at 238, 135 N.E. at 275.

25. *Id.*

26. Many states still adhere to the strict privity requirement, however. See *supra* note 17 and accompanying text.

27. 49 Cal. 2d 647, 320 P.2d 16 (1958).

28. *Id.* at 650, 320 P.2d at 19.

29. 56 Cal. 2d 583, 364 P.2d 685, 15 Cal. Rptr. 821 (1961), *cert. denied*, 368 U.S. 987 (1962).

30. *Id.* at 589, 364 P.2d at 688, 15 Cal. Rptr. at 824.

31. 70 Cal. 2d 223, 449 P.2d 161, 74 Cal. Rptr. 225 (1969).

32. *Id.* at 232-33, 449 P.2d at 167, 74 Cal. Rptr. at 231.

C. Theories of Liability in the Absence of Privity

Courts have recognized two theories for allowing recovery by third parties in legal malpractice actions. Some courts use a "balance of factors" test when determining attorney liability,³³ and others use an intended beneficiary theory of recovery.³⁴ Each theory works as a substitute for privity and gives rise to a duty owed by the attorney.

California has been the forerunner in establishing attorney liability to third parties under the "balance of factors" test. The following factors are considered:

- [1] the extent to which the transaction was intended to affect the plaintiff,
- [2] the foreseeability of harm to him,
- [3] the degree of certainty that the plaintiff suffered injury,
- [4] the closeness of the connection between the defendant's conduct and the injury suffered,
- [5] the moral blame attached to the defendant's conduct,
- and [6] the policy of preventing future harm.³⁵

This test determines whether the lawyer owes a duty to the nonclient and dictates the extension of liability.³⁶ Although other jurisdictions have adopted California's "balance of factors" test, it has been speculated that the test is nothing more than a rule of liability based on third party beneficiary theory.³⁷

Many courts recognize the intended beneficiary exception to privity.³⁸ In rejecting California's balance of factors test, the Pennsylvania Supreme Court in *Guy v. Liederbach*³⁹ expressly adopted the intended beneficiary rule.⁴⁰ In *Guy*, the defendant-attorney was employed by the testator to

33. See *Licata v. Spector*, 26 Conn. Supp. 378, 225 A.2d 28 (1966); *McAbee v. Edwards*, 340 So. 2d 1167 (Fla. Dist. Ct. App. 1976); *Jenkins v. Wheeler*, 69 N.C. App. 140, 316 S.E.2d 354 (1984); *Auric v. Continental Casualty Co.*, 111 Wis. 2d 507, 331 N.W.2d 325 (1983); *Cifu, Expanding Legal Malpractice to Nonclient Third Parties - At What Cost?*, 23 COLUM. J.L. & SOC. PROBS. 1, 11 n.63 (1989) (citing *Franko v. Mitchell*, 158 Ariz. 391, 762 P.2d 1345 (1988)).

34. See, e.g., *Pelham v. Griesheimer*, 92 Ill. 2d 13, 440 N.E.2d 96 (1982); *Guy v. Liederbach*, 501 Pa. 47, 459 A.2d 744 (1983). See also *Davis, Lawyers' Negligence Liability to Nonclients: A Texas Viewpoint*, 14 ST. MARY'S L.J. 405 (1983); Note, *Extending Legal Malpractice Liability to Nonclients - The Washington Supreme Court Considers the Privity Requirement*, 61 WASH. L. REV. 761 (1986).

35. *Heyer v. Flaig*, 70 Cal. 2d 223, 227, 449 P.2d 161, 164, 74 Cal. Rptr. 225, 228 (1969).

36. *Baldock v. Green*, 109 Cal. App. 3d 234, 239, 167 Cal. Rptr. 157, 160 (1980).

37. Note, *The Pelham Decision, Attorney Malpractice and Third Party Nonclient Recovery: The Rise and Fall of Privity*, 3 N. ILL. L. REV. 357 (1983).

38. See *supra* note 34.

39. 501 Pa. 47, 459 A.2d 744 (1983).

40. *Id.* at 59, 459 A.2d at 751.

prepare his will in which the plaintiff was named as a beneficiary. The plaintiff was barred from inheriting under the will, however, because she served as a witness to the document pursuant to the attorney's instructions. The Pennsylvania Court of Common Pleas sustained the attorney's demurrer and dismissed the complaint because of the lack of an attorney-client relationship.⁴¹ In reversing the judgment of the Court of Common Pleas, the Superior Court of Pennsylvania stated, "[t]he better view, we believe, is that, under certain circumstances, an attorney may be liable for damage caused by his negligence to a person intended to be benefitted by his performance irrespective of any lack of privity."⁴² The Supreme Court of Pennsylvania affirmed the judgment of the superior court and noted that the attorney and client must have intended to conduct the transaction for the benefit of the third party in order for him to qualify as an intended beneficiary.⁴³

These two exceptions to the privity requirement enable nonclients to prove the element of duty owed by the attorney when asserting legal malpractice actions. Consequently, nonclients are no longer precluded from bringing such claims. After this extended liability was established, some jurisdictions began to allow nonclients to bring the action *on behalf* of the clients.⁴⁴ In these jurisdictions, clients can assign claims against their attorneys to a third party. Although such an arrangement differs from the exceptions to privity, the practice of assigning claims to nonclients could never have begun under the strict privity rule that only clients have standing to sue. Jurisdictions disagree as to whether legal malpractice claims should be assignable, and much of the debate revolves around the characteristics of the claim.

D. Characteristics of the Legal Malpractice Action

Usually, it is the attorney-client relationship that satisfies the element of duty owed by the attorney in a claim for legal malpractice.⁴⁵ A breach of that duty occurs when an attorney fails "to use such skill, prudence and diligence as lawyers of ordinary skill and capacity commonly possess and exercise in the performance of the tasks which they undertake."⁴⁶

41. *Id.* at 50, 459 A.2d at 746.

42. Guy v. Liederbach, 279 Pa. Super. 543, 546, 421 A.2d 333, 335 (1980), *modified*, 501 Pa. 47, 459 A.2d 744 (1983).

43. Guy Liederbach, 501 Pa. 47, 62, 459 A.2d 744, 752 (1983).

44. See *infra* notes 119-23 and accompanying text.

45. But see *supra* notes 33-43 and accompanying text (identifying two theories that give rise to a duty in the absence of an attorney-client relationship).

46. Lucas v. Hamm, 56 Cal. 2d 583, 591, 364 P.2d 685, 689, 15 Cal. Rptr. 821, 825 (1961).

When such failure proximately causes damage, it gives rise to an action in tort. Since in the usual case, the attorney undertakes to perform his duties pursuant to a contract with the client, the attorney's failure to exercise the requisite skill and care is also a breach of an express or implied term of that contract. . . .⁴⁷

Courts have asserted many different considerations in determining whether legal malpractice is a tort claim or a contract claim. Some jurisdictions focus on the uniquely personal nature of legal services and the highly personal relationship between attorneys and their clients. Consequently, these jurisdictions equate legal malpractice actions with torts involving personal injury or "wrongs done to the person of the injured party or his reputation or feelings."⁴⁸ Other courts focus on the damages suffered and recognize that legal malpractice does not result in personal injury, but concerns purely pecuniary interests.⁴⁹ These courts concentrate on the contractual nature of the action. Additionally, legal malpractice actions survive the death of either party.⁵⁰ Although not determinative, this feature is interesting given the common-law rule that "a cause of action in tort abates on death, but an action for breach of contract survives."⁵¹

The waiver of the attorney-client privilege is one final characteristic of legal malpractice claims that is vital to the issue of assignment:

In the ordinary malpractice action brought by a client, the client may not sue for breach of the attorney's duties and also simultaneously prevent the attorney from defending himself by invoking the [attorney-client] privilege. The holder of the privilege, the client, implicitly waives the privilege by filing such a suit.⁵²

47. Goodley v. Wank & Wank, Inc., 62 Cal. App. 3d 392, 394, 133 Cal. Rptr. 83, 85 (1976).

48. *Id.* at 397, 133 Cal. Rptr. at 87. *See also* Clement v. Prestwich, 114 Ill. App. 3d 479, 480, 448 N.E.2d 1039, 1041 (1983) ("A client's claim for malpractice arises from this personal relationship and is a claim that his attorney has breached a personal duty owed to the client.").

49. *See, e.g.*, Thurston v. Continental Casualty Co., 567 A.2d 922, 923 (Me. 1989); Hedlund Mfg. Co. v. Weiser, Stapler & Spivak, 517 Pa. 522, 536, 539 A.2d 357, 359 (1988).

50. *See, e.g.*, Newman v. Gates, 165 Ind. 171, 72 N.E. 638 (1904); Saltmarsh v. Burmard, 151 Mich. App. 476, 391 N.W.2d 382 (1986); Citizens State Bank v. Shapiro, 575 S.W.2d 375 (Tex. Ct. App. 1978); Nellas v. Loucas, 156 W. Va. 77, 191 S.E.2d 160 (1972).

51. North Chicago St. Ry. v. Ackley, 171 Ill. 100, 105, 49 N.E. 223, 225 (1897). *See also* 1 R. MALLEN & J. SMITH, *LEGAL MALPRACTICE* 366 (3d. ed. 1989).

52. Kracht v. Perrin, Gartland & Doyle, 219 Cal. App. 3d 1019, 1024 n.6, 268 Cal. Rptr. 637, 641 n.6 (1990).

Thus, in defending against a legal malpractice claim, attorneys can reveal details of their interactions with clients, seemingly abrogating the attorney-client privilege. Such revelations may be critical in justifying the conduct of the attorney; therefore, this implied waiver of the attorney-client privilege affords the attorney a fair opportunity to assert a defense.

The characteristics of the legal malpractice action will aid in analyzing whether the claim should be assignable. Before embarking on that discussion, however, a review of the law of assignment is useful.

II. THE LAW OF ASSIGNMENT

A. *Assigning Choses of Action*

According to *Corbin on Contracts*, “[a]ssignments are transfers that substitute ‘a new party as the focus of legal relations’ with respect to the thing assigned.”⁵³ Therefore, the assignment of a chose of action is a transfer of the legal right to assert a claim. Under early common law, a chose of action could not be assigned.⁵⁴ As a result of modern legislation and judicial interpretation, however, “[a]signability . . . is now the rule; nonassignability, the exception.”⁵⁵

The simultaneous expansion of assignment of claims and the demise of the privity requirement created favorable conditions for their combination in the assignment of legal malpractice claims. Although both enable a nonclient to bring the lawsuit,⁵⁶ the two concepts differ, as was indicated by the Supreme Court of Pennsylvania: “[p]rivity is *not* an issue in cases involving an assigned claim because the assignee stands in the shoes of the assignor and does not pursue the cause of action in the assignee’s own right.”⁵⁷ Consequently, the assignee acquires no more rights than were possessed by the assignor.⁵⁸ In determining whether an action is assignable, courts have examined several different considerations.

53. 4 A. CORBIN, *CORBIN ON CONTRACTS* § 861 (1951), quoted in *Essex v. Ryan*, 446 N.E.2d 368, 374 (Ind. Ct. App. 1983).

54. See, e.g., *Goodley v. Wank & Wank, Inc.*, 62 Cal. App. 3d 392, 393, 133 Cal. Rptr. 83, 84 (1976); Annotation, *Assignability of Claim for Legal Malpractice*, 40 A.L.R. 4th 684, 685 (1985).

55. *Goodley*, 62 Cal. App. 3d at 393, 133 Cal. Rptr. at 84 (citations omitted).

56. See generally *Clement v. Prestwich*, 114 Ill. App. 3d 479, 480-81, 448 N.E.2d 1039, 1041 (1983).

57. *Hedlund Mfg. Co. v. Weiser, Stapler & Spivak*, 517 Pa. 522, 525, 539 A.2d 357, 358 (1988) (citing *Gray v. Nationwide Mut. Ins. Co.*, 422 Pa. 500, 223 A.2d 8 (1966)).

58. See, e.g., *Essex v. Ryan*, 446 N.E.2d 368, 374 (Ind. Ct. App. 1983).

B. Factors to Determine if Action is Assignable

When determining whether a particular action is assignable, courts consider factors such as the survivability of the action, the nature of the claim, and public policy. The survivability of the action is relevant because, “[a]s a general rule, actions which are deemed to survive death are assignable under the common law.”⁵⁹ Courts have further established that actions arising from a contract or those arising from a tort to real or personal property survive the death of either party.⁶⁰ Conversely, claims involving torts against a person generally do not survive.⁶¹ Therefore, under the survivability test, contract claims and claims for torts against property are assignable. The test also prescribes that claims for torts against a person are not assignable.

Another approach is to bypass the survivability test and to determine assignability according to the nature of the claim. The result, however, is essentially the same:

[C]auses of action for personal injuries arising out of a tort are not assignable nor are those founded upon wrongs of a purely personal nature such as to the reputation or the feelings of the one injured. Assignable are choses of action arising out of an obligation or breach of contract as are those arising out of a violation of a right of property . . . or a wrong involving injury to personal or real property.⁶²

In addition to the survivability test and the nature of the claim, courts also consider public policy in deciding whether to allow the assignment.⁶³ This additional element permits the court to consider the equities involved in the proposed assignment.⁶⁴ Consequently, when deciding whether a chose of action is assignable, courts look to the

59. *Joos v. Drillock*, 127 Mich. App. 99, 102, 338 N.W.2d 736, 738 (1983). *See also* *North Chicago St. Ry. Co. v. Ackley*, 171 Ill. 100, 108, 49 N.E. 222, 225 (1897); *Christison v. Jones*, 83 Ill. App. 3d 334, 337, 405 N.E.2d 8, 10 (1980).

60. *See, e.g., North Chicago St. Ry. Co.*, 171 Ill. at 105, 49 N.E. at 225.

61. *Id.*

62. *Goodley v. Wank & Wank, Inc.*, 62 Cal. App. 3d 392, 393-94, 133 Cal. Rptr. 83, 85 (1976). *See also* *Oppel v. Empire Mut. Ins. Co.*, 517 F. Supp. 1305, 1307 (S.D.N.Y. 1981); *Hedlund Mfg. Co., Inc. v. Weiser, Stapler & Spivak*, 517 Pa. 522, 525-26, 539 A.2d 357, 358 (1988).

63. *Joos*, 127 Mich. App. at 104, 338 N.W.2d at 739. *See generally Clement v. Prestwich*, 114 Ill. App. 3d 479, 448 N.E.2d 1039 (1983); *Picadilly, Inc. v. Raikos*, 555 N.E.2d 167 (Ind. Ct. App. 1990), *argued*, No. 41A01-8908-CV-311 (Ind. Feb. 20, 1991); *Chaffee v. Smith*, 98 Nev. 222, 645 P.2d 966 (1982); *Collins v. Fitzwater*, 277 Or. 401, 560 P.2d 1074 (1977).

64. For a discussion of the public policy considerations regarding the assignment of legal malpractice claims, see *infra* notes 82-118 and 133-42 and accompanying text.

survivability of the action, the nature of the claim, and public policy concerns. The issue of whether a legal malpractice claim is assignable has only recently received attention.⁶⁵ Jurisdictions addressing the issue are divided on the propriety of such an assignment.

III. ARGUMENTS IN OPPOSITION TO THE ASSIGNMENT OF LEGAL MALPRACTICE CLAIMS

As stated above, actions which are deemed to survive death are generally assignable.⁶⁶ Furthermore, legal malpractice actions survive the death of either party.⁶⁷ Thus, the survivability test supports the theory that legal malpractice actions are assignable. The following jurisdictions have overcome this argument, however, in their assertion that such actions are not assignable: Arizona,⁶⁸ California,⁶⁹ Connecticut,⁷⁰ Florida,⁷¹ Illinois,⁷² Kentucky,⁷³ Michigan,⁷⁴ and Nevada.⁷⁵ Courts in these jurisdictions look beyond this common-law test of survivability and base their decisions on the two remaining factors that determine assignability: the nature of the claim and, more importantly, public policy.⁷⁶

A. *Nature of the Legal Malpractice Claim*

The claim of legal malpractice resembles both a tort claim and a contract claim.⁷⁷ The basis for this disparity has been identified as follows:

65. 1 R. MALLEN & J. SMITH, *LEGAL MALPRACTICE* 368 (3d ed. 1989).

66. See *supra* note 59 and accompanying text.

67. See *supra* note 50 and accompanying text.

68. Schroeder v. Hudgins, 142 Ariz. 395, 690 P.2d 114 (1984).

69. Kracht v. Perrin, Gartland & Doyle, 219 Cal. App. 3d 1019, 268 Cal. Rptr. 637 (1990); Jackson v. Rogers & Wells, 210 Cal. App. 3d 336, 258 Cal. Rptr. 454 (1989); Goodley v. Wank & Wank, Inc., 62 Cal. App. 3d 392, 133 Cal. Rptr. 83 (1976).

70. Continental Casualty Co. v. Pullman, Comley, Bradley & Reeves, 709 F. Supp. 44 (D. Conn. 1989) (supporting the nonassignment of legal malpractice claims in dicta).

71. Mickler v. Aaron, 490 So. 2d 1343 (Fla. Dist. Ct. App. 1986); Washington v. Fireman's Fund Ins. Co., 459 So. 2d 1148 (Fla. Dist. Ct. App. 1984).

72. Brocato v. Prairie State Farmers Ins. Ass'n, 166 Ill. App. 3d 986, 520 N.E.2d 1200 (1988); Clement v. Prestwich, 114 Ill. App. 3d 479, 448 N.E.2d 1039 (1983); Christison v. Jones, 83 Ill. App. 3d 334, 405 N.E.2d 8 (1980).

73. Coffey v. Jefferson County Bd. of Educ., 756 S.W.2d 155 (Ky. Ct. App. 1988).

74. American Employer's Ins. Co. v. Medical Protective Co., 165 Mich. App. 657, 419 N.W.2d 447 (1988); Moorhouse v. Ambassador Ins. Co., 147 Mich. App. 412, 383 N.W.2d 219 (1985); Joos v. Drillock, 127 Mich. App. 88, 338 N.W.2d 736 (1983).

75. Chaffee v. Smith, 98 Nev. 222, 645 P.2d 966 (1982).

76. Goodley v. Wank & Wank, Inc., 62 Cal. App. 3d 392, 133 Cal. Rptr. 83 (1976); Christison, 83 Ill. App. 3d at 334, 405 N.E.2d at 8; Joos, 127 Mich. App. at 99, 338 N.W.2d at 736; Chaffee, 98 Nev. at 222, 645 P.2d at 966.

77. See *supra* notes 48-49 and accompanying text.

[The legal malpractice claim] is primarily a tort action for negligence based upon an attorney's failure to exercise a reasonable degree of skill and care in representing his client. Yet, the duty allegedly breached in such an action arose out of the establishment of the attorney-client relationship by a contract for legal services. As was noted previously, the injuries resulting from legal malpractice are not personal injuries, in the strict sense of injuries to the body, feelings or character of the client. Rather, they are pecuniary injuries to intangible property interests. While focus on these aspects of the malpractice cause of action might indicate placement of it under the class of tort actions for injury to personal property, such placement overlooks the personal nature of the relationship, with attendant duties, that exists between an attorney and client. It is a breach of those duties within the relationship which forms the real basis and substance of the malpractice suit.⁷⁸

Although recognizing the dual nature of the claim, these courts maintain that it more closely resembles a tort action for purposes of assignment. For example, an appellate court in Illinois stated, “[t]he cause of action based upon legal malpractice is, in its essence, a tort action for negligence premised upon breach of the attorney's duties to his client.”⁷⁹ Similarly, a California Court of Appeals noted, “the gravamen of [a legal malpractice action] is the negligent breach by defendants of a duty to plaintiff's assignor.”⁸⁰

The nature of the claim is only one reason these courts prohibit assignability. As one court stated, “the assignability of a cause of action must be based upon an analysis of the claim sought to be assigned as well as upon the public policy considerations involved.”⁸¹ In fact, public policy arguments are the primary focus in these jurisdictions.⁸²

78. *Christison v. Jones*, 83 Ill. App. 3d 334, 338, 405 N.E.2d 8, 10 (1980) (citation omitted). See also *Schroeder v. Hudgins*, 142 Ariz. 395, 399, 690 P.2d 114, 118 (1984); *Kracht v. Perrin, Gartland & Doyle*, 219 Cal. App. 3d 1019, 1024, 268 Cal. Rptr. 637, 640 (1990); *Jackson v. Rogers & Wells*, 210 Cal. App. 3d 336, 342, 258 Cal. Rptr. 454, 457 (1989); *Goodley*, 62 Cal. App. 3d at 397, 133 Cal. Rptr. at 87; *Joos*, 127 Mich. App. at 105-06, 338 N.W.2d at 739.

79. *Christison*, 83 Ill. App. 3d at 336, 405 N.E.2d at 9.

80. *Goodley*, 62 Cal. App. 3d at 395, 133 Cal. Rptr. at 86.

81. *Joos v. Drillock*, 127 Mich. App. 99, 104, 338 N.W.2d 736, 739 (1983).

82. See generally *Goodley v. Wank & Wank, Inc.*, 62 Cal. App. 3d 392, 133 Cal. Rptr. 83 (1976); *Washington v. Fireman's Fund Ins. Co.*, 459 So. 2d 1148 (Fla. Dist. Ct. App. 1984); *Clement v. Prestwich*, 114 Ill. App. 3d 479, 448 N.E.2d 1039 (1983); *Christison*, 83 Ill. App. 3d at 334, 405 N.E.2d at 8; *Moorhouse v. Ambassador Ins. Co.*, 147 Mich. App. 412, 383 N.W.2d 219 (1985); *Joos*, 127 Mich. App. at 99, 338 N.W.2d at 736.

B. Public Policy Considerations

Courts opposing the assignment of legal malpractice claims assert several public policy concerns in support of their position including: (1) the sanctity of the attorney-client relationship; (2) the preservation of the attorney-client privilege; (3) the potential for commercialization; (4) the risk of collusion; and (5) the illogical arguments asserted by the assignee.

1. *Sanctity of the Attorney-Client Relationship.*—The relationship between an attorney and a client has been the primary focus of courts opposing assignment. These courts cite the “uniquely personal nature of legal services and the contract out of which a highly personal and confidential attorney-client relationship arises.”⁸³ This concern was apparent in *Goodley v. Wank & Wank, Inc.*⁸⁴ In *Goodley*, the assignor, Eleanor Katz, was represented by the law firm of Wank & Wank in a divorce proceeding. Wank & Wank negligently advised Katz that she need not give her husband’s life insurance policies to them for safe-keeping or obtain a court order precluding her husband from changing the terms. Her husband subsequently found and cancelled the policies. Shortly thereafter, he died. Katz then assigned her legal malpractice action against Wank & Wank to Harry Goodley.⁸⁵ In sustaining the defendants’ motion for summary judgment, the California Court of Appeals recognized the importance of the attorney-client relationship, stating, “[t]he relation between attorney and client is a fiduciary relation of the very highest character, and binds the attorney to most conscientious fidelity. . . .”⁸⁶

The Illinois Court of Appeals in *Christison v. Jones*⁸⁷ asserted the same sentiment. Christison was a trustee in bankruptcy and claimed ownership to any cause of action for legal malpractice which the bankrupt might have against the attorney, Jones. Jones was allegedly negligent in his representation of the bankrupt in litigation prior to the bankruptcy. In ruling against the assignability of the legal malpractice claim, the *Christison* court cited public policy concerns regarding the fiduciary relationship between the attorney and the client and stated, “[t]he relationship is a confidential one to be highly honored and guarded by

83. *Joos*, 127 Mich. App. at 103, 338 N.W.2d at 738 (quoting *Goodley*, 62 Cal. App. 3d at 395, 133 Cal. Rptr. at 86). See also *Jackson v. Rogers & Wells*, 210 Cal. App. 3d 340, 342, 258 Cal. Rptr. 454, 457 (1989).

84. 62 Cal. App. 3d 389, 133 Cal. Rptr. 83 (1976).

85. *Id.* It is unclear from the opinion why Katz assigned her claim to Harry Goodley.

86. *Id.* at 395, 133 Cal. Rptr. at 86 (quoting *Cox v. Delmas*, 99 Cal. 104, 123, 33 P. 836, 839 (1893)).

87. 83 Ill. App. 3d 334, 405 N.E.2d 8 (1980).

the attorney.''⁸⁸ Courts have also illustrated the uniquely personal nature of the attorney-client relationship by noting that legal services performed by an attorney are not delegable without the client's prior consent.⁸⁹ In reality, the confidential and highly personal nature of this relationship has been of paramount concern when courts disallow assignment.⁹⁰

The apparent looseness of a particular attorney-client relationship may be irrelevant. An appellate court in Michigan expressly recognized this irrelevancy by refusing to examine the closeness of the particular attorney-client relationship at issue.⁹¹ According to the court in *Moorhouse v. Ambassador Insurance Co.*,⁹² such an approach would lead "to the impossible task of dissecting the closeness of an attorney-client relationship in evaluating the validity of every assignment of a cause of action for legal malpractice."⁹³

Some courts state that the decision to assert a legal malpractice claim is uniquely within the discretion of the client.⁹⁴ This principle also underlies the decision of the Supreme Court of Nevada in *Chaffee v. Smith*.⁹⁵ In *Chaffee*, the assigned claim had not yet been asserted by the client-assignor. The court disallowed the assignment stating, "[t]he decision as to whether to bring a malpractice action against an attorney is one peculiarly vested in the client."⁹⁶ The court expressly reserved its opinion regarding the assignment of a previously asserted legal malpractice claim.⁹⁷ The preservation of the attorney-client privilege is another public policy consideration used by some courts that disallow assignment.

2. *Preservation of the Attorney-Client Privilege*.—Clients who bring legal malpractice actions implicitly waive the right to assert the attorney-

88. *Id.* at 338, 405 N.E.2d at 10-11.

89. See, e.g., *Goodley v. Wank & Wank, Inc.*, 62 Cal. App. 3d 389, 396, 133 Cal. Rptr. 83, 86 (1976); *Christison*, 83 Ill. App. 3d at 338, 405 N.E.2d at 11.

90. See generally *Schroeder v. Hudgins*, 142 Ariz. 395, 399, 690 P.2d 114, 118 (1984); *Mickler v. Aaron*, 490 So. 2d 1343, 1344 (Fla. Dist. Ct. App. 1986); *Washington v. Fireman's Fund Ins. Co.*, 459 So. 2d 1148, 1149 (Fla. Dist. Ct. App. 1984); *Brocato v. Prairie State Farmers Ins. Ass'n*, 166 Ill. App. 3d 986, 989, 520 N.E.2d 1200, 1201 (1988); *Moorhouse v. Ambassador Ins. Co.*, 147 Mich. App. 412, 414, 383 N.W.2d 219, 221 (1985) ("the entire *Joos* decision hinged on just such a close, personal relationship"); *Joos v. Drillock*, 127 Mich. App. 99, 105, 338 N.W.2d 736, 739 (1983).

91. *Moorhouse v. Ambassador Ins. Co.*, 147 Mich. App. 412, 383 N.W.2d 219 (1985).

92. *Id.*

93. *Id.* at 414, 383 N.W.2d at 221.

94. See, e.g., *Christison v. Jones*, 83 Ill. App. 3d 334, 339, 405 N.E.2d 8, 11 (1980).

95. 98 Nev. 222, 645 P.2d 966 (1982).

96. *Id.* at 224, 645 P.2d at 966.

97. *Id.*

client privilege.⁹⁸ This privilege is not waived, however, when someone other than the client files the suit.⁹⁹ Consequently, the attorney-client privilege is preserved when the malpractice claim is involuntarily assigned.

For example, in *Kracht v. Perrin, Gartland & Doyle*,¹⁰⁰ Brenda Kracht, the nonclient-assignee, filed a complaint against Charles Hogue, the client-assignor. Hogue retained the services of the law firm Perrin, Gartland & Doyle to defend him against Kracht's lawsuit. During the course of the proceedings, the opposing attorneys failed to adequately respond to discovery requests made by Kracht. As a result of this failure to respond, judgment was entered in favor of Kracht.¹⁰¹

Kracht contended that the deficient discovery responses by Hogue were proximately caused by the negligence of Hogue's attorneys and that if the attorneys had exercised proper skill and care in their representation, the judgment in favor of Kracht would not have been entered. Kracht subsequently sought and won a court order compelling Hogue to assign the choses of action he held against his attorneys to Kracht. Kracht then filed a claim for legal malpractice against Hogue's attorneys. Because Hogue did not bring the legal malpractice action, his attorneys were bound by their privileged relationship with him. In ruling against the assignment of the claim, the *Kracht* court focused on the involuntary nature of the assignment: “[a]n involuntary assignment thus unfairly prejudices either the attorney (by precluding any defense based on privileged communications) or the client (by permitting the assignee to waive the privilege without the client's consent).”¹⁰² In addition to considering the importance of the attorney-client relationship and privilege, courts cite the problems of potential commercialization as a rationale for opposing the assignment of legal malpractice actions.

3. *Potential for Commercialization*.—The court in *Goodley v. Wank & Wank, Inc.*¹⁰³ identified potential abuses due to assignment:

The assignment of such claims could relegate the legal malpractice action to the market place and convert it to a commodity to be exploited and transferred to economic bidders who have never had a professional relationship with the attorney and to whom the attorney has never owed a legal duty, and who have never had any prior connection with the assignor or his rights. The

98. See *supra* note 52 and accompanying text.

99. *Kracht v. Perrin, Gartland & Doyle*, 219 Cal. App. 3d 1019, 1024 n.6, 268 Cal. Rptr. 637, 641 n.6 (1990).

100. 219 Cal. App. 3d 1019, 268 Cal. Rptr. 637 (1990).

101. *Id.* at 1021, 268 Cal. Rptr. at 638.

102. *Id.* at 1024 n.6, 268 Cal. Rptr. at 641 n.6.

103. 62 Cal. App. 3d 389, 133 Cal. Rptr. 83 (1976).

commercial aspect of assignability of choses in action arising out of legal malpractice is rife with probabilities that could only debase the legal profession. The almost certain end result of merchandizing such causes of action is the lucrative business of factoring malpractice claims which would encourage unjustified lawsuits against members of the legal profession, generate an increase in legal malpractice litigation, promote champerty and force attorneys to defend themselves against strangers. The endless complications and litigious intricacies arising out of such commercial activities would place an undue burden on not only the legal profession but the already overburdened judicial system, restrict the availability of competent legal services, embarrass the attorney-client relationship and imperil the sanctity of the highly confidential and fiduciary relationship existing between attorney and client.¹⁰⁴

According to the *Goodley* court and the other courts that have cited this language with approval,¹⁰⁵ legal malpractice actions will become commercialized if they are made assignable. Malpractice claims will become a commodity to be bought and sold to the highest bidder, malpractice litigation will increase, and attorneys will be forced to defend against strangers.¹⁰⁶ Furthermore, there will be a decrease in the availability of legal services because attorneys will be more selective in accepting new clients. Assignment will thereby render a disservice to both the profession and the public.¹⁰⁷ Malpractice premiums might also rise.¹⁰⁸ In addition to commercializing legal malpractice actions, assignments may encourage collusion.

4. *Risk of Collusion*.—At least one court has disallowed the assignment of legal malpractice claims because of the risk of collusion.¹⁰⁹ In *Coffey v. Jefferson County Board of Education*,¹¹⁰ the client-assignor was a defendant in a negligence action brought by the nonclient-assignee.

104. *Id.* at 397, 133 Cal. Rptr. at 87.

105. *Continental Casualty Co. v. Pullman, Comley, Bradley & Reeves*, 709 F. Supp. 44, 50-51 (D. Conn. 1989); *Jackson v. Rogers & Wells*, 210 Cal. App. 3d 340, 343, 258 Cal. Rptr. 454, 458 (1989); *Brocato v. Prairie State Farmers Ins. Ass'n*, 166 Ill. App. 3d 986, 989, 520 N.E.2d 1200, 1202 (1988); *Clement v. Prestwich*, 114 Ill. App. 3d 479, 481, 448 N.E.2d 1039, 1041-42 (1983); *Joos v. Drillock*, 127 Mich. App. 99, 103, 338 N.W.2d 736, 738 (1983).

106. *Goodley*, 62 Cal. App. 3d at 397, 133 Cal. Rptr. at 87.

107. *Id.*

108. *Kracht v. Perrin, Gartland & Doyle*, 219 Cal. App. 3d 1019, 1023, 268 Cal. Rptr. 637, 640 (1990); *Jackson*, 210 Cal. App. 3d at 348, 258 Cal. Rptr. at 461.

109. *Coffey v. Jefferson County Bd. of Educ.*, 756 S.W.2d 155 (Ky. Ct. App. 1988).

110. *Id.*

On the day of the trial, the client-assignor appeared before the court and confessed a judgment of \$1,000,000. At the same time, he attempted to assign all claims he had for legal malpractice against his former attorneys to the nonclient-assignee. Because the court did not enter a judgment in the original action, it found that there was no proof that the client-assignor suffered actual damage from the alleged malpractice.¹¹¹ The court went on to say, “[i]n addition, it appears to us that this transaction is so collusive that same should be held to be against public policy.”¹¹²

As *Coffey* illustrates, the nonclient-assignee and the client-assignor may enter into a collusive agreement whereby the client admits liability or agrees to settle, in exchange for which the nonclient accepts the assignment of a legal malpractice action as satisfaction of that judgment. Relief from the costs and inconveniences of defending a lawsuit gives the client-assignor an incentive to enter the agreement, even in the absence of actual negligence by the attorney. Because the client-assignor may be insolvent or have limited assets, the nonclient-assignee will have a similar incentive to enter the agreement. Consequently, because the assignor and the assignee may engage in collusion when assigning the action, assignment has been held to violate public policy.¹¹³

5. *Illogical Arguments Asserted by Assignee.*—The context of the proposed assignment is often one in which the nonclient-assignee has brought an action and has prevailed against the client-assignor.¹¹⁴ In these situations, the assignor’s action for legal malpractice involves alleged negligence by the attorney in defending the client in the action brought by the nonclient.¹¹⁵ Such an assignment produces illogical and contradictory arguments by the nonclient-assignee in the subsequent legal malpractice lawsuit. These arguments may directly contradict those advanced in the original lawsuit against the client-assignor. As the California Court of Appeals stated in *Kracht*:

[A] malpractice suit filed by the former adversary is “fraught with illogic” and unseemly arguments: In the former lawsuit [the nonclient-assignee] judicially averred and prove she was

111. *Id.* at 156-57.

112. *Id.* at 157.

113. *Id.*

114. See, e.g., *Kracht v. Perrin, Gartland & Doyle*, 219 Cal. App. 3d 1019, 268 Cal. Rptr. 637 (1990); *Jackson v. Rogers & Wells*, 210 Cal. App. 3d 340, 258 Cal. Rptr. 455 (1989); *Clement v. Prestwich*, 114 Ill. App. 3d 479, 448 N.E.2d 1039 (1983); *Picadilly, Inc. v. Raikos*, 555 N.E.2d 167 (Ind. Ct. App. 1990); *Moorhouse v. Ambassador Ins. Co.*, 147 Mich. App. 412, 383 N.W.2d 219 (1985); *Chaffee v. Smith*, 98 Nev. 222, 645 P.2d 966 (1982).

115. See *supra* note 114.

entitled to recover against [the client-assignor]; but in the legal malpractice lawsuit [the nonclient-assignee] must judicially aver that, but for attorney's negligence, she *was not entitled* to have recovered against [the client-assignor]. Reduced to its essence, [the nonclient-assignee's] argument in the malpractice action is "To the extent I was not entitled to recover, I am now entitled to recover."¹¹⁶

This problem was also recognized in the concurring opinion in *Picadilly, Inc. v. Raikos*.¹¹⁷ This opinion noted that to allow such a collateral attack against a judgment may lead to a practice whereby defeated defendants routinely assign legal malpractice claims as a means of satisfying the judgments against them.¹¹⁸ Illogical arguments and collateral attacks against judgments will be eliminated if the assignment of legal malpractice claims is not allowed.

The above considerations are used to varying degrees by courts that oppose assignability, but the courts all agree that public policy prohibits such a practice. These arguments are found unpersuasive, however, by those courts that allow such assignments.

IV. ARGUMENTS IN SUPPORT OF THE ASSIGNMENT OF LEGAL MALPRACTICE CLAIMS

Courts consider three factors in determining whether a chose of action is assignable: the survivability of the action, the nature of the claim, and public policy concerns. Although the test of survivability supports the assignment of legal malpractice claims, it is rarely cited by those courts allowing such assignments. Instead, courts ruling in favor of the assignability of the actions have focused on the nature of the claim and public policy considerations. The following states have allowed the assignment of legal malpractice actions: Indiana,¹¹⁹ Maine,¹²⁰ New York,¹²¹ Oregon,¹²² and Pennsylvania.¹²³

116. *Kracht*, 219 Cal. App. 3d at 1025, 268 Cal. Rptr. at 641 (emphasis in original).

117. 555 N.E.2d 167, 171 (Ind. Ct. App. 1990) (Baker, J., concurring), *argued*, No. 41A01-8908-CV-311 (Ind. Feb. 20, 1991).

118. *Id.*

119. *Id.* at 167.

120. *Thurston v. Continental Casualty Co.*, 567 A.2d 922 (Me. 1989).

121. *Oppel v. Empire Mut. Ins. Co.*, 517 F. Supp. 1305 (S.D.N.Y 1981); *American Hemisphere Marine Agencies, Inc. v. Kreis*, 40 Misc. 2d 1090, 244 N.Y.S.2d 602 (1963).

122. *Collins v. Fitzwater*, 277 Or. 401, 560 P.2d 1074 (1977).

123. *Hedlund Mfg. Co. v. Weiser, Stapler & Spivak*, 517 Pa. 522, 539 A.2d 357 (1988).

A. Nature of the Legal Malpractice Claim

Most jurisdictions favoring assignability cite the nature of the claim in their analyses.¹²⁴ These courts note the economic, rather than personal, nature of the harm alleged.¹²⁵ The court in *Hedlund Manufacturing Co. v. Weiser, Stapler & Spivak*¹²⁶ expounded upon this theory. The dispute arose after Mervin Martin sold his business to Hedlund Manufacturing Company. Along with his business, Martin turned over his rights to use and license a pending patent he owned on a manure spreader that he invented and manufactured. After the sale was complete, the parties learned that the patent was denied because Martin's attorneys neglected to file the application on time. Subsequently, Martin assigned his cause of action for legal malpractice to Hedlund Manufacturing.

In deciding whether to allow the assignment, the *Hedlund* court recognized that "the court must determine whether the claim is for damages or personal injury. This inquiry is critical to the viability of the assigned cause of action in that we do not permit the assignment of a cause of action to recover for personal injuries."¹²⁷ The court concluded that the assignment was valid because the damages resulting from legal malpractice were pecuniary in nature.¹²⁸ The claim asserted more closely resembled an action involving property rights than one for personal rights.¹²⁹

Similar reasoning was used in *Oppel v. Empire Mutual Insurance Co.*¹³⁰ In *Oppel*, the assignee's young son was seriously injured when he was struck by the assignor's car. The nonclient-assignee offered to settle for the \$10,000 limit on the client-assignor's car insurance policy with Empire Mutual Insurance Company. Empire Mutual expressed a willingness to settle for about \$500 less than the policy limit. The nonclient-assignee alleged that the settlement amounted to bad faith and thus, brought an action against the client-assignor resulting in an award of \$420,850 in damages.¹³¹

Subsequently, the client-assignor transferred his malpractice actions against Empire Mutual and his attorney to the nonclient-assignee. In approving the assignment, the court noted, "[h]ere there is no allegation

124. See, e.g., *Oppel*, 517 F. Supp. at 1307; *Thurston*, 567 A.2d at 923; *American Hemisphere*, 40 Misc. 2d at 1092, 244 N.Y.S.2d at 604; *Hedlund Mfg. Co.*, 517 Pa. at 525, 539 A.2d at 358-59.

125. See, e.g., *Thurston v. Continental Casualty Co.*, 567 A.2d 922, 923 (Me. 1989).

126. 517 Pa. 522, 539 A.2d 357 (1988).

127. *Id.* at 525, 539 A.2d at 358.

128. *Id.* at 526, 539 A.2d at 359.

129. *Id.*

130. 517 F. Supp. 1305 (S.D.N.Y. 1981).

131. *Id.* at 1306.

that the attorney's acts caused any personal injury, only pecuniary. This claim is also assignable."¹³² Jurisdictions allowing the assignment of legal malpractice actions also justify their conclusions by citing public policy considerations.

B. Public Policy Considerations

The public policy considerations articulated by jurisdictions allowing the assignment of legal malpractice claims differ from those used by the states opposing it. Efficiency and equity are the chief policy considerations expressed by courts that allow assignment.

1. *Efficiency*.—The Supreme Court of Maine in *Thurston v. Continental Casualty Co.*¹³³ recognized efficiency as an advantage of assigning legal malpractice claims. In *Thurston*, the nonclient-assignee brought a products liability action against 3K Kamper Ko., the client-assignor. Inadequate legal representation and misconduct on the part of 3K's attorneys allegedly caused the nonclient-assignee to be awarded an amount in excess of 3K's insurance policy limit. Because 3K could not pay the excess judgment, it agreed to assign all its choses of action against its attorney to the nonclient-assignee.

In rejecting an argument regarding the sanctity of the attorney-client relationship, the court stated:

The argument that legal services are personal and involve confidential attorney-client relationships does not justify preventing a client like 3K from realizing the value of its malpractice claim in what may be the most efficient way possible, namely, its assignment to someone else with a clear interest in the claim who also has the time, energy and resources to bring the suit.¹³⁴

The court was persuaded that the expenditures required to maintain a malpractice action should be incurred by the party most able to afford them.¹³⁵ The court also considered the equities involved before ruling in favor of the proposed assignment.¹³⁶

2. *Equity*.—The *Thurston* court dismissed the concern that legal malpractice actions would become commercialized¹³⁷ by noting that the nonclient-assignee had an "intimate connection with the underlying lawsuit."¹³⁸ The *Picadilly* court similarly disregarded this concern by noting

132. *Id.* at 1307.

133. 567 A.2d 922 (Me. 1989).

134. *Id.*, quoted in *Picadilly, Inc. v Raikos*, 555 N.E.2d 167 (Ind. Ct. App. 1990).

135. *Id.* at 923.

133. *Id.*

137. See *supra* notes 103-08 and accompanying text.

138. *Thurston v. Continental Casualty, Co.*, 567 A.2d 922, 923 (Me. 1989).

that the client-assignor asserted the legal malpractice claim before assigning it and that the discovery process had already factually developed the case.¹³⁹

In addition to rejecting the public policy concerns professed by other jurisdictions about the potential for commercialization, these jurisdictions considered equitable policy concerns of their own. For example, in *Collins v. Fitzwater*,¹⁴⁰ an attorney negligently drafted the interest-bearing promissory notes issued by a corporation for which the client-assignor served as a director. It was later determined that the notes qualified as unregistered securities. Consequently, the client-assignor was held liable under the Blue Sky Law to the purchasers of these unregistered securities. As a means of avoiding bankruptcy, the client-assignor negotiated covenants not to execute them with the purchasers, in exchange for an assignment of the legal malpractice action against his attorney.

In allowing the assignment of the action, the *Collins* court focused on the inequities of the situation:

As a matter of necessity, laymen who act as corporate directors must often rely upon the expertise and diligence of corporate counsel when intricate legal questions are at issue. When a corporate attorney errs in the performance of his legal duties, we can think of no reason why the laymen rather than the attorney should bear the ultimate burden of the error. Therefore, we conclude that public policy does not prohibit a nonculpable director from seeking indemnification from a culpable attorney through a suit for legal malpractice, or otherwise. Similarly, since [the client-assignor] had a valid claim against the [attorney], we believe that [the nonclient-assignees'] assignment gave them an enforceable right of action, and that that assignment is not void as against public policy.¹⁴¹

This equitable sentiment was also expressed by the court in *Hedlund*:

We will not allow the concept of the attorney-client relationship to be used as a shield by an attorney to protect him or her from the consequences of legal malpractice. Where the attorney has caused harm to his or her client, there is no relationship that remains to be protected.¹⁴²

139. Picadilly, Inc. v. Raikos, 555 N.E.2d 167, 169 (Ind. Ct. App. 1990), *argued*, No. 41A01-8908-CV-311 (Ind. Feb. 20, 1991).

140. 277 Or. 401, 560 P.2d 1074 (1977).

141. *Id.* at 406, 560 P.2d at 1078.

142. *Hedlund Mfg. Co. v. Weiser, Stapler & Spivak*, 517 Pa. 522, 526, 539 A.2d 357, 359 (1988).

These courts allow the assignment of legal malpractice claims as a means of holding attorneys accountable for their malpractice. These public policy concerns, the advantage of efficiency, and the inherent nature of the claim, constitute the reasons asserted by those jurisdictions in the minority that allow legal malpractice actions to be assigned.

V. SUGGESTED RESOLUTION: ASSIGNMENT OF JUDGMENTS

Courts deciding against the assignability of legal malpractice actions, as well as those deciding in favor of it, have identified compelling arguments in support of their respective positions. A viable compromise is to assign the resulting *judgments* of legal malpractice actions, rather than assigning the *chooses of action*.

A. *Background on the Assignment of Judgments*

Under early common law, judgments could not be assigned.¹⁴³ As with the assignment of *chooses of action*,¹⁴⁴ however, this rule was often modified by legislation and judicial interpretation.¹⁴⁵ Indiana Code section 34-1-31-1 provides an example of legislation that establishes the propriety of the assignment of judgments:

Vesting of title in assignee: Judgments and decrees of a court of record for the recovery of money may be assigned by the plaintiff or complainant. The assignees successively on or attached to the entry of the judgment or decree and the assignment, when attested by the clerk of the court, or vests the title to the judgment or decree in each assignee successively.¹⁴⁶

The assignee succeeds to the ownership of the judgment and to all of the rights of the assignor therein.¹⁴⁷ Statutory provisions, such as Indiana Code section 34-1-31-4, enable assignees to assert actions on the judgment in their own names.¹⁴⁸ Similarly, the assignee takes the judgment subject to the defenses that could have been used against the judgment when it was owned by the assignor.¹⁴⁹ The assignor is divested of all rights and interests in the judgment.¹⁵⁰

143. *Baker v. Wood*, 157 U.S. 212 (1895).

144. See *supra* notes 54-55 and accompanying text.

145. See, e.g., *Boyd v. Sloan*, 335 Mo. 163, 71 S.W.2d 1065 (1934).

146. IND. CODE § 34-1-31-1 (1988).

147. See, e.g., *Moorman v. Wood*, 117 Ind. 144, 19 N.E. 739 (1888).

148. IND. CODE § 34-1-31-4 (1988) ("Any action which the plaintiff or complainant in such judgment or decree might have thereon, may be maintained in the name of the assignee."').

149. See, e.g., *Frankel v. Garrard*, 160 Ind. 209, 66 N.E. 687 (1902).

150. *Moorman*, 117 Ind. at 144, 19 N.E. at 739.

A judgment may be assigned any time after it is entered in the trial court, even if an appeal is pending.¹⁵¹ It is not enforceable, however, until it is final.¹⁵² An assigned *judgment* would be effective even in those jurisdictions that have ruled against the assignability of legal malpractice *claims* because “[e]ven though a cause of action is one which is not deemed to be assignable, a final judgment into which such cause of action in tort is merged may be assigned.”¹⁵³

The assignability of a legal malpractice judgment was recently addressed in Michigan in *Weston v. Dowty*.¹⁵⁴ In *Weston*, the Michigan Court of Appeals allowed the assignment of the *judgment* even though Michigan consistently rules against the assignability of legal malpractice *chooses of action*.¹⁵⁵ In *Weston*, the assignor was a defendant in a personal injury lawsuit in which he was represented by attorneys Dowty and Schlussel. Because the attorneys failed to comply with discovery orders, a default judgment was entered against the client-assignor. Before a trial was held on the issue of damages, the client-assignor and the plaintiffs to the personal injury action entered into a consent judgment in the amount of \$200,000. In the consent judgment, the client-assignor agreed to file a legal malpractice claim against his attorneys and to give any monies awarded to him in a judgment to the nonclient-assignees, less costs and attorney fees. The attorneys filed a motion for summary judgment in the legal malpractice action, asserting that the consent judgment was the assignment of a legal malpractice claim and was, therefore, invalid. The court disagreed with the attorneys and upheld the assignment of the judgment pursuant to the terms of the consent judgment.¹⁵⁶

The *Weston* court acknowledged that *Joos v. Drillock*¹⁵⁷ and *Moorhouse v. Ambassador Insurance Co.*¹⁵⁸ prohibited the assignment of legal malpractice actions.¹⁵⁹ It distinguished the case at bar stating, “In the instant case, [the client-assignor] did not assign the claim or cause of action to [the nonclient-assignee]. [The client-assignor] merely agreed to give [the nonclient-assignee] any proceeds recovered.”¹⁶⁰ Additionally,

151. See, e.g., *Bias v. Ohio Farmers Indem. Co.*, 28 Cal. App. 2d 14, 81 P.2d 1057 (1938).

152. See, e.g., *Pacific Gas & Elec. Co. v. Nakano*, 12 Cal. 2d 711, 87 P.2d 700 (1939).

153. 46 AM. JUR. 2D *Judgments* § 884 (1969).

154. 163 Mich. App. 238, 414 N.W.2d 165 (1987).

155. See *supra* note 74.

156. *Weston*, 163 Mich. App. at 241, 414 N.W.2d at 166.

157. 127 Mich. App. 99, 338 N.W.2d 736 (1983).

158. 147 Mich. App. 412, 383 N.W.2d 219 (1985).

159. *Weston v. Dowty*, 163 Mich. App. 238, 241-42, 414 N.W.2d 165, 166-67 (1987).

160. *Id.* at 242, 414 N.W.2d at 167.

the client-assignor maintained the legal malpractice lawsuit himself. For these reasons, the court found that it was not bound by the holdings in *Joos* and *Moorhouse*.¹⁶¹ Because the right to assert the action was vested in the client-assignor, the assignment of the judgment was upheld.¹⁶²

The result in *Weston v. Dowty* illustrates a viable compromise to the issue of whether legal malpractice actions should be assignable. Courts have analyzed three factors in deciding the assignability of particular actions: the survivability of the action, the nature of the claim, and public policy considerations. Currently, little emphasis is placed on the survivability of the action. The nature of the claim is also nondeterminative because legal malpractice claims resemble both negligence claims and actions for breach of contract.¹⁶³ Thus, courts have difficulty labeling the action as either a tort or contract claim. Because of these deficiencies, most decisions are based on public policy considerations. Michigan prohibits the assignment of legal malpractice actions and allows the assignment of legal malpractice judgments.¹⁶⁴ This approach addresses the public policy concerns advanced by both courts that oppose the assignability of the claim and those that support it.

B. *Public Policy Considerations in Opposition to the Assignment of the Action*

1. *Sanctity of the Attorney-Client Relationship*.—The importance of the attorney-client relationship is the primary reason the court disallowed the assignment of the legal malpractice claim in *Goodley v. Wank & Wank, Inc.*¹⁶⁵ The attorneys in *Goodley* breached their fiduciary duty when they gave the client-assignor erroneous advice. The court ruled against the assignment of the claim because the personal nature of the attorney-client relationship gave rise to the duty that was breached.¹⁶⁶ Because the attorneys did not owe a duty to the nonclient-assignee, the nonclient was not allowed to bring the action.¹⁶⁷ If the judgment is assigned, however, the client-assignor will maintain the malpractice claim through its final disposition. The attorneys will then be defending against

161. *Id.*

162. *Id.*

163. See *supra* notes 48-49 and accompanying text.

164. See *supra* note 74 (legal malpractice claims are not assignable). But see *Weston v. Dowty*, 163 Mich. App. 238, 414 N.W.2d 165 (1987) (legal malpractice judgments are assignable).

165. 62 Cal. App. 3d 389, 133 Cal. Rptr. 83 (1976).

166. *Id.* at 395, 133 Cal. Rptr. at 86.

167. *Id.*

the party to whom they owed the fiduciary duty, rather than defending against strangers.

Another advantage to the assignment of the judgment is its consistency with the common-law rule that only those in privity with the attorney can sue for legal malpractice.¹⁶⁸ Unlike *Goodley*, in which a nonclient maintained the action, this proposed resolution ensures that only a client will assert a malpractice claim.¹⁶⁹ Furthermore, the decision whether to pursue the action will remain with the client. Assigning the judgment satisfies these objectives because nothing is transferred until a judgment is entered in favor of the client-assignor. Therefore, this suggested resolution respects the importance of the attorney-client relationship.

2. *Preservation of the Attorney-Client Privilege*.—The court in *Kracht v. Perrin, Gartland & Doyle*¹⁷⁰ held that legal malpractice claims may not be involuntarily assigned because the attorney-client privilege has not been waived.¹⁷¹ In *Kracht*, the nonclient-assignee sought a court order compelling the client to assign all actions against his attorneys to the nonclient-assignee. In ruling against the assignability of the claim, the court focused on preserving the attorney-client privilege. Had the assignment been upheld, the attorney would have been bound by the privilege and unable to put forth a fair defense.

Assigning the judgment will alleviate this policy concern. Because the malpractice action against the attorney will be brought by the client-assignor, the attorney-client privilege will be impliedly waived. For example, in *Kracht*, the attorney could have used previously privileged information to assert a fair defense against his former client. Because the attorney-client privilege is not preserved when the judgment is assigned, this policy consideration will not be at issue.

3. *Potential for Commercialization*.—Assigning a judgment does not generate the same potential abuses that may result from assigning the action. The inherent risks in the assignment of the claim were identified by the court in *Goodley v. Wank & Wank, Inc.*¹⁷² The court in *Goodley* was concerned that assignments will create an economic market in which legal malpractice actions will serve as commodities.¹⁷³ According to the

168. See *supra* note 17 and accompanying text.

169. A nonclient may be able to assert such a claim, however, if the theory falls into one of the two exceptions to privity: the "balance of factors" test or the third party beneficiary theory. See *supra* notes 33-43 and accompanying text.

170. 219 Cal. App. 3d 1019, 268 Cal. Rptr. 637 (1990).

171. *Id.* at 1025, 268 Cal. Rptr. at 641.

172. 62 Cal. App. 3d 389, 133 Cal. Rptr. 83 (1976).

173. *Id.* at 397, 133 Cal. Rptr. at 87.

court, such a market could result in an increase in unjustified lawsuits and a decrease in available legal services.¹⁷⁴

An assignment of the judgment does not take place until the judgment is final. The nonclient-assignee will agree to the assignment only if the final judgment is in favor of the client-assignor. Prevailing in the legal malpractice action is a prerequisite for the client-assignor to be able to use the assignment to satisfy a debt owed to the nonclient. Consequently, the malpractice action will be instituted only if the client-assignor has a fair chance of prevailing. Unlike the assignment of the claim, assigning the judgment places the risk of losing the malpractice lawsuit on the client-assignor. This decreases both the number of assignments and the number of malpractice lawsuits litigated. Therefore, the overburdened judicial system will be afforded some relief if the judgment, rather than the action, is assigned. Unjustified lawsuits will also decrease because the client-assignor will not bring the malpractice suit unless the chances of prevailing are favorable.

Legal malpractice claims will not become a commodity to be sold to the highest bidder because only clients will be able to bring the actions. Furthermore, the decrease in the number of lawsuits will mean that attorneys will not have to be more selective in accepting clients than they would be if no assignment was allowed. Thus, the availability of legal services will not diminish. A rise in malpractice insurance will not be as inevitable as it would be if actions were assigned. For all of these reasons, assigning the judgment from a legal malpractice action results in fewer potential abuses than assigning the action itself.

4. *Risk of Collusion.*—The court in *Coffey v. Jefferson County Board of Education*¹⁷⁵ disallowed the assignment of a legal malpractice claim because of the risk that the client-assignor and the nonclient-assignee would enter a collusive agreement to the detriment of the attorney.¹⁷⁶ In *Coffey*, the client was the defendant in a personal injury action brought by the nonclient. The assignment of the legal malpractice claim was part of a pretrial settlement between the client-assignor and the nonclient-assignee. In these situations, the danger is that the client will offer to settle for a high amount and then assign the legal malpractice claim in satisfaction of that debt. The client-assignor will agree to the assignment because it will satisfy the obligation owed, and the nonclient-assignee will agree to the assignment because such a large amount may not be recoverable from the client-assignor. Therefore, a collusive agreement may be entered even if the attorney was not negligent.

174. *Id.*

175. 756 S.W.2d 155 (Ky. Ct. App. 1988).

176. *Id.* at 157.

This type of collusive agreement will not be possible if the judgment is assigned. Under this suggested resolution, an assignment cannot take place until the attorney's liability is established by a final judgment. Consequently, there is no risk that the parties to the assignment will enter a collusive agreement that will result in a lawsuit against a faultless attorney. An assignment may occur only if the court finds that the attorney committed legal malpractice. Any subsequent agreement to assign the judgment will not be collusive because the client-assignor will incur actual damages as a result of the attorney's negligence. Therefore, the assignment of the judgment removes the incentive and opportunity for collusion.

5. *Illogical Arguments Asserted by the Assignee.*—Some courts recognize the illogical arguments asserted by the nonclient-assignee as a reason for disallowing the assignment of legal malpractice claims. These courts disapprove of malpractice claims in which the nonclient-assignee argues that the attorneys are liable because the nonclient should not have prevailed in the original action against the client-assignor.

Because the nonclient-assignee would not be maintaining the legal malpractice action if the judgment was assigned, these contradictory arguments would not be asserted. Furthermore, the arguments asserted by the client-assignor will be consistent with those made in the original action against the nonclient-assignee. For example, if the judgment was assignable in *Picadilly, Inc. v. Raikos*,¹⁷⁷ Picadilly, as the client-assignee, would be the plaintiff throughout the entire trial against its attorneys. Picadilly's contention would be that it should have prevailed in the original action brought by Colvin and that it lost because of its attorneys' inadequate representation. Picadilly would assert its nonculpability regarding the dram shop claim in both the original action and the malpractice action.

C. Public Policy Considerations in Support of the Assignment of the Action

1. *Efficiency.*—One reason for assigning the action is efficiency.¹⁷⁸ Because the nonclient-assignee may have more time, energy, and resources to bring the suit, it should be assignable. This argument was advanced by the court in *Thurston v. Continental Casualty Co.*¹⁷⁹ In *Thurston*, the client-assignor was a corporation that had suffered financial hardships making its ability to institute a lawsuit tenuous. The efficiency argument

177. 555 N.E.2d 167, 169 (Ind. Ct. App. 1990), *argued*, No. 41A01-8908-CV-311 (Ind. Feb. 20, 1991).

178. See *supra* note 134 and accompanying text.

179. 567 A.2d 922 (Me. 1989).

recognizes the costs that the client-assignor must incur in maintaining a legal malpractice claim against its attorneys.

This efficiency objective could be equally satisfied under this proposed resolution. As in *Weston v. Dowty*, the nonclient-assignee could be assigned the judgment, less costs and attorney fees.¹⁸⁰ This procedure financially enables clients to assert legal malpractice actions, provided they feel their claim has merit. Furthermore, the assignment can be attached to the entry of the judgment, so that it is effective as soon as the judgment is final.¹⁸¹ Therefore, efficiency will be as well-served with the assignment of the judgment as with the assignment of the action.

2. *Equity*.—Some courts allow the assignment of legal malpractice actions citing the objective that culpable attorneys should be held liable for their negligence.¹⁸² For example, in *Hedlund Manufacturing Co. v. Weiser, Stapler & Spivak*,¹⁸³ one reason that the court allowed the assignment of the claim was to ensure that the attorney would bear the cost of failing to file the patent application on time.¹⁸⁴ Permitting the client-assignor to transfer the legal malpractice action to the nonclient-assignee was apparently regarded as a means of realizing this end.

Assigning the judgment, instead of the action, also serves this objective. Although the plaintiff to the action will be the client-assignor rather than the nonclient-assignee, attorneys will still be made to pay for their misconduct. Additionally, the nonclient-assignee will not be left without a means of recovery under this suggested resolution. Recovery will be to the same extent as if the nonclient-assignee brought the malpractice action. Thus, the equitable concerns maintained by the courts that allow the assignment of claims will be favorably addressed if the judgments are assigned.

Although it has not been addressed by the courts, another equitable goal will be achieved under this suggested resolution. As between the client-assignor and the nonclient-assignee, it is equitable that the former bear the risk of losing the legal malpractice action. When the malpractice action is transferred by the client-assignor, it is often done to satisfy a previous judgment owed to the nonclient-assignee.¹⁸⁵ Therefore, if the court in the legal malpractice action finds in favor of the attorney, the nonclient-assignee is left without compensation.

180. *Weston v. Dowty*, 163 Mich. App. 238, 239-40, 414 N.W.2d 165, 166 (1987).

181. For a statutory example, see *supra* text accompanying note 146.

182. See *supra* notes 141-42 and accompanying text.

183. 517 Pa. 522, 539 A.2d 357 (1988).

184. *Id.* at 526, 539 A.2d at 359.

185. See *supra* note 114 and accompanying text.

Conversely, when the *judgment* of the malpractice action is assigned, it is the client-assignor who bears the risk of losing. Because the assignment of the judgment is not effective until it is final, it will transpire only if the client-assignor prevails against the attorney. Consequently, there will be nothing to assign if the court determines that malpractice was not committed. Such a decision would establish that it was the client-assignor's wrongful conduct, and not the attorney's representation, that led to the disposition of the original suit in favor of the nonclient-assignee. In this situation, it is equitable that the client-assignor will be unable to use assignment of the action to relieve the obligation owed to the nonclient-assignee.

An equitable argument for assigning the judgment can also be made when there was no prior lawsuit between the client and nonclient. In *Hedlund*, neither the client-seller of the company nor the nonclient-buyer acted wrongfully. The attorney had the sole responsibility for filing the patent application. Nevertheless, equity requires that the client-assignor bear the risk of losing the legal malpractice action against his attorney. The client selected the attorney to handle the affairs of the business and should therefore bear the consequences of this choice.

D. Possible Disadvantages to Assigning the Judgment

One potential problem with assigning the judgment is that the client-assignor is forced to remain in the legal malpractice action for its duration. This differs from the assignment of the claim in which the client-assignor is relieved of all liability before the trial begins. Because the lawsuit against the attorney could take years before its final disposition, the client-assignor bears an additional burden. Meanwhile, the nonclient-assignee must wait to receive the money that is owed him by the client.

On the other hand, it is equitable for the client-assignor to bear the risk of losing the malpractice action. Thus, the client should be the one who brings the claim, no matter the duration. Furthermore, the release of liability afforded the client-assignor when the claim is transferred creates an incentive to assign unjustified claims against the attorney. As for the disadvantage to the nonclient-assignee of having to wait for the compensation, this is not as significant after examining the realities of the alternative.

A nonclient-assignee will probably not accept *any* assignment unless it is the only means of recovery. If the client-assignor is able to satisfy the obligation to the nonclient-assignee without an assignment, then the nonclient will insist on such reimbursement. Consequently, the assignment usually takes place because it is the best means of recovery for the assignee. Because someone must prevail against the attorney before the nonclient-

assignee can be compensated, a trial is unavoidable. The length of the trial will be no greater if the client-assignor maintains the action than if the nonclient-assignee maintains it. Therefore, this disadvantage to the assignment of the judgment is not persuasive.

Furthermore, in the event that the attorney prevails in the malpractice action, the nonclient-assignee will still own the outstanding debt owed by the client-assignor under this proposed resolution. If the client subsequently acquires assets, the nonclient will still have a legal right to the sum that was awarded. Such protection of the nonclient will not be available if the unfounded malpractice action is assigned because the client's obligation will be satisfied before the malpractice trial begins.

Another potential disadvantage of assigning the judgment is that the nonclient-assignee may have more of an incentive to pursue the malpractice claim than the client-assignor because the latter will be seeking a recovery for someone else. This contention is unfounded, however, because the nonclient-assignee will have an incentive to prevail even though the proceeds will go to the nonclient. If the client-assignor is able to effectuate an assignment of the judgment, the client's obligation to the nonclient is discharged. Any assets obtained by the client-assignor thereafter will not be subject to the debt previously owed to the nonclient-assignee.

In addition, it may be alleged that the nonclient-assignee will have an incentive to recover a greater amount, while the client-assignor will try to obtain the amount owed to the nonclient. Although differing objectives from the assignor and assignee may be inherent, they will probably not result in greatly different awards. The amount of damages in the legal malpractice lawsuit will be the economic harm incurred by the client which was proximately caused by the attorney's negligence.¹⁸⁶ Consequently, no matter how great the nonclient-assignee's incentive to get more, the amount of the recovery will be limited. For example, in *Picadilly, Inc. v. Raikos*, a judgment of \$150,000 was entered against Picadilly in the dram shop action brought by the nonclient-assignee. The theory in the subsequent legal malpractice action against Picadilly's attorneys was that the attorneys proximately caused an erroneous judgment to be entered against Picadilly. The damages sought would be equal to that incurred by Picadilly. Therefore, regardless of whether the malpractice action was brought by Picadilly or by the nonclient-assignee, the amount of the recovery would be limited to \$150,000. Accordingly, the greater incentive of the nonclient-assignee would probably not result in a significantly higher award.

Although assigning the judgment of a legal malpractice action may result in certain disadvantages, these are negligible when compared to the

186. See, e.g., *Pickens, Barnes & Abernathy v. Heasley*, 328 N.W.2d 524 (Iowa 1983); *Collins ex rel. Collins v. Perrin*, 108 N.M. 714, 778 P.2d 912 (1989); *Rizzo v. Haines*, 520 Pa. 484, 555 A.2d 58 (1989); *Cosgrove v. Grimes*, 774 S.W.2d 662 (Tex. 1989).

problems inherent in assigning the action. Assigning the judgment is preferable because it furthers the public policy concerns advanced by the courts on both sides of the issue. It addresses both the advantages and disadvantages that accompany the assignment of legal malpractice actions. The Indiana Court of Appeals in *Picadilly* could have encouraged such a resolution by first denying Charles Colvin, the nonclient-assignee, the right to sue Picadilly's attorneys. The court could then have established in dicta that any judgment obtained against the attorneys by Picadilly could have been properly assigned to Colvin after it became final. If Indiana and other jurisdictions were to encourage the assignment of judgments, they would be acknowledging the dual nature of the claim as well as competing public policy considerations.

VI. CONCLUSION

For approximately the last thirty years, statutes and judicial decisions have suppressed the strict rule that privity is required in legal malpractice actions. As an extension of this expanded liability, courts began to decide whether legal malpractice actions should be assignable. The majority of the courts that have spoken on this issue have held that public policy prohibits such assignment. The following policy concerns have been advanced in opposition to the assignability of the claims: the sanctity of the attorney-client relationship, the preservation of the attorney-client privilege, the potential for commercialization, the risk of collusion, and the illogical arguments asserted by the assignee. The minority view allows the assignment of legal malpractice actions. These courts maintain that the claim resembles a contract action, which is assignable at law. The courts that support the assignment of legal malpractice actions also maintain that two public policy considerations are furthered: efficiency and equity.

The court of appeals in Michigan has established a precedent that favorably addresses the concerns advanced by both the majority and the minority jurisdictions on this issue. Michigan prohibits the assignment of legal malpractice actions, but allows the resulting judgments from such actions to be assigned. This type of resolution favorably addresses the public policy considerations that have been advanced by both sides. Furthermore, the assignment of the judgment puts the risk of losing the legal malpractice action on the client-assignor rather than on the nonclient-assignee. This arrangement results in an equitable outcome. Potential disadvantages to this type of assignment are relatively insignificant when compared to the assignment of the chose of action. Courts in other jurisdictions should follow Michigan's approach of allowing the assignment of the legal malpractice judgment because it is a viable compromise to the present controversy.

Indiana Law Review

Volume 25

1991

Number 1

BOOK REVIEW

Narrative and the Legal Neighborhood

THOMAS PATRICK GANNON*

JAY A. ZIEMER**

NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW. Edited with introductory comments and discussion questions by *David Ray Papke*. Liverpool, U.K.: Deborah Charles Publications, 1990. Pp. ix, 368. \$35.00.

In his prefatory remarks to *Narrative and the Legal Discourse*, Professor David Ray Papke¹ employs an appealing metaphor to describe the tentative relationship between storytelling and the dominant legal order: “[n]arrative lives across the street from law’s established home.”² If the articles collected in this volume are any indication, then the law does indeed have some noisy new neighbors. But demographic changes are not always easy. One can already imagine the grumblings coming from law’s home — “there goes the neighborhood.”

Just who are these boisterous, Simpsonesque intruders? And why might their presence cause discomfort to the stodgy, old law family across the street? With respect to the latter question, an appreciation of narrative casts doubt on fundamental assumptions about the law in much the same

* Assistant Director, The Center for Law and Health, Indiana University School of Law-Indianapolis; B.A., 1987, Indiana University; J.D., 1990, Indiana University School of Law-Indianapolis.

** Assistant Public Defender, Cook County Public Defender’s Office, Chicago; B.A., 1987, Marquette University; J.D., 1991, Indiana University School of Law-Indianapolis.

1. Professor of Law and American Studies, Indiana University School of Law-Indianapolis; A.B., 1964, Harvard College; M.A., 1978, Yale University; J.D., 1973, Yale University; Ph.D., 1984, The University of Michigan.

2. D. PAPKE, NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 3 (1990).

manner that the popular television show, *The Simpsons*, undermines America's mythical *Leave-It-To-Beaver* notions of family life. What once seemed "natural," now appears quaint. A paradox of storytelling, however, is that the capacity to form alternative narratives, especially alternative legal narratives, is often limited by the stock of stories already told.

As for the identity of these newcomers, *Narrative and the Legal Discourse* provides at least a provisional answer. Using a group of influential, sometimes provocative, law-related articles published within the last half decade, Professor Papke has compiled an accessible and up-to-date anthology exploring the connections between law and narrative in four broad areas: legal education, litigation, legal doctrine (namely, the appellate opinion), and alternative legal narratives. The majority of contributors teach at law schools around the United States. One author is a practicing litigator although others teach sociology, criminal justice, and anthropology.

Why narrative? One reason is that storytelling is the basic, predominant form of human communication. Stories are found in all cultures and all cultural products, including law. These articles persistently remind the reader of the numerous ways in which storytelling is used to comprehend, shape, and organize our understanding of law and the social order. From our first days in law school through our professional lives as attorneys, judges, and scholars, we use language as the "tool" of our "craft." Indeed, such metaphors reveal an underlying confidence in the precision of language.

Yet, since at least the beginning of the twentieth century, philosophers, linguists, literary theorists, and others have recognized that the relationship between language, meaning, and truth is highly problematical. The focus of inquiry has shifted to the complex processes of interpretation. From the various humanistic and semiotic perspectives³ found in *Narrative and the Legal Discourse*, legal communication is seen as a subjective or interpretive process mediated by social, cultural, or psychological factors through which meaning, and thus "reality," are not so much received as generated or constructed. If narrative, as Papke defines it, is "the transformation of events and sentiments into stories which impart meaning,"⁴ then narrative has even greater importance for legal study because it challenges the notion that law is fixed, or that one can neatly separate theories, facts, and values.

Despite the centrality of narrative as a mode of social and legal communication, Professor Papke argues that legal thinkers in the United

3. Legal semiotics is the study of signs and codes that underlie legal informational exchanges. See Tiefenbrun, *Legal Semiotics*, 5 CARDOZO ARTS & ENT. L.J. 89, 96 (1986).

4. D. PAPKE, *supra* note 2, at 4.

States have either ignored or displayed hostility to narrative theory because jurisprudential thought has been largely dominated by legal scientism or legal positivism.⁵ Like many disciplines, modern jurisprudence and legal education were shaped by the postenlightenment, scientific milieu of the nineteenth century. Borrowing from the methods of the natural sciences, Dean Langdell of the Harvard Law School established the fundamentals of legal formalism: the law library was to become the "laboratory" in which case law would be analyzed.⁶ Using the Socratic method and syllogism, law professors guide students to the "correct" answer through deductive logic.⁷

Papke also suggests that law's isolation as a discipline has obscured the inherent subjectivity of legal content. Although various movements such as legal realism have managed to broach the walls of the citadel from time to time, law has remained largely isolated from the interdisciplinary developments that affect other fields. In the last twenty years, however, law and economics, critical legal studies, feminist legal theory, and the law and literature movement have contributed to make the study of law and narrative possible.

Taken together, the articles in *Narrative and the Legal Discourse* constitute an emerging alternative jurisprudence which may serve to integrate the legal neighborhood. Papke argues that narrative sensitivity might not only have a beneficial effect on the smug insularity of legal scholarship, but that it also encourages everyone, including the public,

to conceptualize the law differently. [The law] ceases to be limited, settled and formal and becomes instead fluid, contested and even

5. The famous English scholar H.L.A. Hart proffered five useful definitions of legal positivism:

- (1) the contention that laws are commands of human beings.
- (2) the contention that there is no necessary connection between law and morals or law as it is and ought to be.
- (3) the contention that the analysis (or study of the meaning) of legal concepts is (a) worth pursuing and (b) to be distinguished from historical inquiries into the causes or origins of laws, from sociological inquiries into the relation of law and other social phenomena, and from the criticism or appraisal of law whether in terms of morals, social aims, "functions," or otherwise.
- (4) the contention that a legal system is a "closed logical system" in which correct legal decisions can be deduced by logical means from predetermined legal rules without reference to social aims, policies, moral standards, and
- (5) the contention that moral judgments cannot be established or defended as statements of fact can, by rational argument, evidence, or proof.

Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 601-02 n.25 (1958).

6. D. PAPKE, *supra* note 2, at 2.

7. For a discussion of the development of this method, see R. STEVENS, LAW SCHOOL: LEGAL EDUCATION IN AMERICA FROM THE 1850s TO THE 1980s 51-64 (1983).

contradicted. . . . [N]arrative enables us to break free of assumptions that the law is objectively created and applied and to appreciate instead the subjectivity of the law. We can more freely enjoy and participate in the ongoing process of re-creating the law. . . . [W]e can understand law not as restriction and control but rather as a realm of possibilities.⁸

Nevertheless, law and narrative scholars may disagree among themselves about the purpose and goals of their study.

One can already detect different perspectives among the various contributors to *Narrative and the Legal Discourse*. Papke himself discerns two strains within the law and narrative movement: a humanistic strain and a more analytical semiotic strain, dedicated to a methodologically rigorous, even scientific, study of signs.⁹ Indeed, some readers may find that the more analytical articles on litigation in Part Two are rather slow-going and less interesting than other sections.

Bernard Jackson, Queen Victoria Professor of Law at the University of Liverpool, is by far the most articulate and theoretical contributor to Part Two. Although a full appreciation of his essay presumes some familiarity with poststructural theory, Jackson's thesis is essentially that narrative is an inherent feature of legal reasoning: neither legal rules nor facts exist autonomously as the uncomplicated subject matter of deductive logic or syllogistic reasoning. They are instead, Jackson argues, narrative constructs applied by a set of legal actors, each with his or her own experiences.¹⁰ Although these important insights should not be overlooked, some of the articles offer little more than sophisticated quantifying, strikingly devoid of normative conclusions about how such information might be used once learned. Indeed, Papke hints that the more humanistic members of the law and narrative household would be disturbed by their kinfolks' "tendency toward a new legal science, one capable of the denial which [the humanists find] so troubling in the dominant legal culture."¹¹

As recent law school graduates, the authors of this review found the articles on legal education relevant to their own experiences. For example, Professor James Elkins of the West Virginia University College of Law summarily observes that "[a]n education in law is a love/hate relationship."¹² Drawing on the journals he assigned first year law students

8. D. PAPKE, *supra* note 2, at 5.

9. See *id.* at 3-4.

10. Jackson, *Narrative Models in Legal Proof*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 159, 170-71 (D. Papke ed. 1990).

11. D. PAPKE, *supra* note 2, at 4.

12. Elkins, *The Quest for Meaning: Narrative Account of Legal Education*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 10, 11 (D. Papke ed. 1990).

to keep,¹³ Elkins charts the changes and disillusionment many of his students experienced.

Elkins argues that severe time constraints, an overemphasis on grades, dull and repetitious exercises, and anti-intellectualism in the classroom, contribute to the loss of student idealism. In many ways, law school is a hierarchical and alienating place. These characteristics are undoubtedly attributable in part to the stultifying Langdellian educational model described above, as well as to a larger cultural belief reflected in the law wherein emotion is devalued and virtually everything is treated as a commodity.

The other two essays in the section on legal education offer intriguing ways in which legal pedagogy might be made more humane and rewarding. Professor Andrew McThenia, Jr. of the Washington and Lee University School of Law argues that teaching is a form of storytelling, a transaction between teller and audience that can result in a mutual learning experience. Implicit in McThenia's argument is the recognition that the public and private spheres overlap, that these are in some sense arbitrary distinctions:

Whoever the storyteller is, she is shaped by one or more narratives. And she is carried by the stream of her own history to that moment of meeting an audience which itself brings one or more stories. As these streams merge, none of the individual lives can ever be the same again.¹⁴

McThenia says that meaning is collectively shaped and is contingent in the sense that one can never be sure of "the truth."

Because of his epistemological skepticism, his insistence that the teaching, learning, and practice of law are value-laden and his belief that teaching is a form of mutual storytelling, McThenia concludes that to be an effective teacher he must reveal who he is: a Christian, an academic, and the loved one of an alcoholic. He says "[t]he issue is *how* to do it, and not *whether* to do it."¹⁵ McThenia's argument suggests that the Langdellian emphasis on objectivity is constantly distorted by values and bias, especially in the classroom.

For David O. Friedrichs, Professor of Sociology and Criminology at the University of Scranton, legal understanding has an undeniable experiential dimension, a sentiment reflecting Oliver Wendell Holmes's observation that "[t]he life of the law has been not logic: it has been experience."¹⁶ Friedrichs argues that narratives occupy a marginal, sub-

13. He actually gave them this option instead of taking exams or writing papers!

14. McThenia, *Telling a Story About Storytelling*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 30, 31 (D. Papke ed. 1990).

15. *Id.* at 37 (emphasis in original).

16. See O.W. HOLMES, THE COMMON LAW 1 (1881).

versive, even “heretical” status in relation to both the orthodoxy of legal education and practice, as well as to undergraduate education, because narratives lack “efficiency” for a professional culture slavishly obsessed with time and reason and are negatively associated with children’s stories.¹⁷ He claims that the use of parables, poems, literature, biography, and autobiography in the classroom can lead students to appreciate the normative, subjective aspects of law and thus enable them to challenge established doctrines.

The articles in Part Three also speak to legal education by focusing on the problems of legal doctrine in the United States that are typically exemplified by case law. Applying literary theory to three of Shakespeare’s plays, Professor John Denvir of the University of San Francisco School of Law argues that the traditional “statist,” positivistic approach of jurisprudence in this country contributes to the continuing problems of racism and other constitutional issues, especially with respect to fashioning remedies.

Borrowing from literary theorist Northrop Frye, Denvir advances what he calls a “jurisprudence of comedy” that embraces “the twin goals of liberation and reconciliation.”¹⁸ He states that:

[H]uman passion, not legal rules, must be the starting place for a theory of law: the major task of jurisprudence must be to create legal institutions that transform human passion into a force for social regeneration, rather than to spin out justifications for sterile rules and often violent repression.¹⁹

He argues that an appreciation of literature counterbalances the black-letter emphasis of positivism in legal education. Law’s supposed neutrality, abstraction, and scientism stand in contrast to literature’s subjectivity, complexity, and passion.

Denvir’s jurisprudence of comedy also has an overtly communitarian political dimension which “recognize[s] that moral norms are usually created at the community level, rather than [by] the state.”²⁰ Thus, he disdains the philosophical assumptions of liberals and others who exalt the individual against the state in matters of constitutional law. He adopts, instead, a communitarian or group-oriented approach which emphasizes “the relationships between competing moral and legal visions within that

17. Friedrichs, *Narrative Jurisprudence and Other Heresies: Legal Education at the Margin*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 43, 45 (D. Papke ed. 1990).

18. Denvir, *William Shakespeare and the Jurisprudence of Comedy*, in NARRATIVE AND THE LEGAL DISCOURSE 183, 184 (D. Papke ed. 1990).

19. *Id.* at 183.

20. *Id.* at 197.

state . . . and underscore[s] the necessity for the state to mediate through its law these insular legal universes and thereby mend tears in the larger social fabric.”²¹ The proper constitutional role of the state, he asserts, is to protect less powerful groups and to upset the status quo.²²

As an example, Denvir examines J. Anthony Lukas’s study of public school desegregation in Boston. Denvir argues that Judge Arthur Garrity’s busing decree was a “disaster” for many black families, a form of scapegoating, a “betrayal” for white working-class families who could not afford to send their children to private schools and was hypocritical essentially because it permitted privileged white liberals to assuage their guilt without directly confronting the status quo.²³ Denvir suggests that the traditional approach to constitutional law perpetuates racism in three ways: by failing to address “private” acts of discrimination as beyond the pale of “state action,” by distinguishing between *de jure* and *de facto* acts, and by distinguishing constitutional liability from remedy.²⁴ Whether or not one agrees with Denvir’s philosophical assumptions, his jurisprudence of comedy offers a refreshing perspective on continuing sociolegal problems. His article also raises a perplexing problem: “the moral imperative to act without the consolation of certainty that our acts are morally justified.”²⁵

By far, the most intriguing and provocative articles in *Narrative and the Legal Discourse* are found in Part Four. These readings address the strategies and difficulties of constructing alternative legal narratives. Drawing on the memoirs of nineteenth century property criminals, Professor Papke contends that these accounts paradoxically reflect the dominant culture’s attitudes: “Rather than speaking as critical outsiders, the criminal memoirists of the period proffer autobiographical narratives which champion the institutions, norms and values most valorized by legitimate Americans. The criminal memoirs, grounded in a strong sense of professionalism, illustrate the power of societal hegemony.”²⁶ Papke’s interpretation suggests that law and legal attitudes are imbued with ideology—not simply “propaganda,” but a normative vision about how the world works, social relations, and one’s place within that order. In this sense, ideology has a mythical aspect. Ideologies are powerfully constructed and reinforced by society and seen as “natural” or unalterable.

21. *Id.*

22. *Id.*

23. *Id.* at 200.

24. *Id.* at 201.

25. *Id.* at 198-99.

26. Papke, *Legitimate Illegitimacy: The Memoirs of Nineteenth-Century Professional Criminals*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 226, 226 (D. Papke ed. 1990).

To challenge this “natural” order, one must first confront one’s social conditioning or the stories one has already learned.

“Success” in altering one’s own consciousness, however, does not ensure that others will accept these views. Jim Thomas, Professor of Sociology at Northern Illinois University, clarifies this point in an article examining the petitions of prisoners for postconviction relief.²⁷ Professor Thomas observes that legal storytelling is a complex linguistic, rhetorical, and conceptual process involving both storyteller and audience.²⁸ Although storytellers differ in their narrative ability to frame issues in patterned and predictable legalistic frames, Thomas shows that how law clerks and judges interpret stories is important in determining whether a specific claim will be successful or even heard. Interpretation is thus ineluctably linked with notions of justice.

The most interesting article in *Narrative and the Legal Discourse* was written by Marie Ashe, Professor of Law at the West Virginia University School of Law.²⁹ What distinguishes Ashe’s piece from the others is not only her subject matter, but her method. Using narrative and literary techniques, she offers profound intellectual insights while engaging the reader with her story. Ashe’s article is narrative as praxis.

Professor Ashe begins by describing her anger at the legal objectification of women:

Whenever I read law relating to women and motherhood, I find myself sickened. When I read *Roe v. Wade* I am filled with anger; when I read the *Baby M* trial court decision, I am enraged. When I hear women referred to as “surrogates,” I have the same reaction as arises when I hear women called “bitches” or “sluts.” Feelings of humiliation, of indignation, of desperation, of horror, of rage. Reading *A.C.*, I feel something close to despair.

Often, in the last several weeks, I have set aside my notes and readings concerning motherhood and law. I leave them with a sense of hopelessness. Often I have picked up some needlework—sewing, embroidery, needlepoint, knitting—seeking some respite from the feelings that overwhelm me, restoration. The rhythm of my fingers becomes a rhythm of my inner being, a peace in my breast. A dropped stitch. A gentle flutter. A minor

27. Thomas, *Prisoner Cases as Narrative*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 237 (D. Papke ed. 1990).

28. See *id.* at 239-44.

29. Ashe, *Zig-Zag Stitching and the Seamless Web: Thoughts on “Reproduction” and the Law*, in NARRATIVE AND THE LEGAL DISCOURSE: A READER IN STORYTELLING AND THE LAW 262 (D. Papke ed. 1990).

interruption of rhythm and pattern. I pick it up easily, drawing it into the larger design. I exist in a silent space. Untroubled.

Law reaches every silent space. It invades the secrecy of women's wombs. It breaks every silence, uttering itself. Law-language, jurisdiction. It defines. It commands. It forces.

Law as the seamless web we believe and die in. I cannot think of a single case involving legal regulation of motherhood without thinking of all. They constitute an interconnected network of variegated threads.³⁰

Ashe appropriates and reinterprets the famous legal metaphor of the seamless web³¹ to show how the institutionalization of medicine and law have used language to abstract, reduce, simplify, define, and categorize women.

By "deconstructing"³² this metaphor, Ashe also evokes the various cultural myths associated with traditional or devalued women's work: the Miller's Daughter, the Lady of Shalott, Penelope, Arachne, Ariadne, and Medea.³³ She then reconstructs the metaphor by linking it to a personal story about her grandmother's quilt and by urging women to define themselves through a new kind of writing and thinking "[m]arked by our varying rhythms and cycles. Our stitches will seldom be straight. Zig-zag stitchings and zig-zag thought. Useful (as in buttonholing) for definition; (as in edging seams) for strength; (as in embroidering) for beauty."³⁴ This metaphor then serves to define the structure of her piece. Indeed, Ashe is such a fine writer that she has perhaps created an entirely new, literary legal form.

By discussing her personal experience with birth, abortion, and miscarriages, Ashe constructs an ethics of "reproduction." She says, "[i]t seems to me that the departure point for such exploration must be women's own accounts of our experiences, uttered with a commitment of faithfulness to the truths of female bodies suppressed in the dominant discourse."³⁵ She rejects generalization, recognizing that each woman must

30. *Id.*

31. While the image of law as a seamless web has frequently been used to describe the interconnectedness of all areas of common law, its initial appearance may be attributable to the English legal historian Frederic William Maitland. "Such is the unity of all history that any one who endeavours to tell a piece of it must feel that his first sentence tears a seamless web." Maitland, *Prologue to a History of English Law*, 14 L.Q. REV. 13, 13 (1898).

32. Deconstructionism is a method or technique used to interpret text which destabilizes "common sense" meanings. See Tiefenbrun, *supra* note 3, at 148-50.

33. Ashe, *supra* note 29, at 263.

34. *Id.* at 264.

35. *Id.* at 380.

determine this for herself and thus, exposes the fallacy of essentialism.³⁶ She argues that law and medicine are guilty of this error:

The medicalization of abortion—like that of childbirth and pregnancy—has set women at a distance from the blood ceremonies of our bodies, placing us at the mercy of a technological priesthood that denies the sacred, detaching us from the physicality and the cultural implications of violence and bloodshed. We have become the victims, and not the agents, of bloodshed.³⁷

Ashe's article is a powerful normative statement which is bound to provoke reaction. More than any other article in this volume, it reveals the possibilities that narrative study can contribute to the legal discourse.

Narrative and the Legal Discourse is a stimulating and suggestive introduction to the emerging study of narrative jurisprudence. As such, the book raises more questions than it answers. Readers interested in more in-depth treatment of the various intellectual and philosophical issues raised by the book, however, will find ample resource material in the footnotes and discussion questions. The success of *Narrative and the Legal Discourse* lies in its critique of what Papke calls “dominant legal conceptualizations.”

Narrative and the Legal Discourse challenges the dominant view that science is the only legitimate form of inquiry. Narrative theory undermines the positivistic premise that “facts” stand alone or that they have some philosophical primacy. Narrative theory shifts the debate to how meaning is constructed and challenges the notion that the law is determinate or that rules can simply be plucked from case law and applied neutrally. In this respect, narrative theory raises questions about one of our most cherished beliefs—the Rule of Law. If law is an essentially interpretive process, how can there by anything but a rule of “men”?

On another level, narrative jurisprudence offers a powerful tool for those seeking to change or reinforce the existing legal system. The diverse perspectives of the contributors to *Narrative and the Legal Discourse* suggest that narrative jurisprudence has only just begun to establish itself in the legal neighborhood.

36. “[T]he notion that every class or group of things has an essential or fundamental nature, common to every member of the class, and that the process of defining consists in isolating and identifying this common nature or intrinsic property.” L. LLOYD & M. FREEMAN, *LLOYD'S INTRODUCTION TO JURISPRUDENCE* 53 (5th ed. 1985).

37. Ashe, *supra* note 29, at 282.

